Accident Analysis and Prevention 45 (2012) 507–516

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data

Kun-Feng Wu*, Paul P. Jovanis

Civil and Environmental Engineering, Larson Institute, Penn State, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 13 June 2011
Received in revised form 8 August 2011
Accepted 1 September 2011

Keywords:
Traffic safety
Crash surrogate
Naturalistic driving study data analysis

ABSTRACT

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 paper connects the original research on traffic conflicts to the contemporary literature concerning crash surrogates using the crash-to-surrogate ratio, π. 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. The paper concludes with a summary of empirical results and suggestions for future research.

1. Introduction

1.1. 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 relationship (Hauer, 1982; Hauer and Gårder, 1986; Davis et al., 2008; Tarko et al., 2009): *Number of crashes expected to occur on an entity during a certain period of time* (λ) = *number of crash surrogates occurring on an entity in that time* (c)* *crash-to-surrogate ratio for that entity* (π). Mathematically,

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

Here we take the liberty of using the more current term, "surrogate", instead of the original term "conflict". Eq. (1) must hold for any meaningful crash surrogate. And hence, Eq. (1) is the conceptual foundation for this crash surrogate analysis. Eq. (1) provides a definitional link between the expected number of crashes, λ; the number of observed surrogate events, c; and the crash-conflict conversion factor, π. If one can develop a method to estimate π 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 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 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 paper then tests the conceptualization empirically, followed by conclusions and lessons learned for future research.

1.2. Traffic 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

* Corresponding author. Tel.: +1 8143211183.
E-mail addresses: kxw930@psu.edu (K.-F. Wu), ppj2@engr.psu.edu (P.P. Jovanis).

0001-4575/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2011.09.002

Table 1
Hydén (1987) conversion factor table.

	Car–car crash/conflict	Car–bicycle crash/conflict	Car–pedestrian crash/conflict
Traffic class 1: All situations in low speed intersections and situations in high speed intersections with only turning cars involved	(x_1, y_1)	(x_2, y_2)	(x_3, y_3)
Traffic class 2: Situations in signalized intersections with only turning cars involved	(x_4, y_4)	(x_5, y_5)	$(x_6\ y_6)$
Traffic class 3: Situations in high speed intersections with at least one straight forward going car involved	(x_7, y_7)	(x_8, y_8)	(x_9, y_9)
Traffic class 4: Situations in signalized intersections with at least one straight forward going car involved	(x_{10}, y_{10})	(x_{11}, y_{11})	(x_{12}, y_{12})

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 Hyden, 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 s 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 π in Eq. (1)), which represents the relationship between conflicts and crashes (Hydén, 1987). First, crashes and conflicts are split with regard to the variables "traffic class" and "kind of road users", as shown in Table 1. "Traffic class" can be seen as a driving context variable, and "kind of road user" can be seen as event type variables. Oppe (1986) suggested a similar classification using maneuver type and severity level.

There are 12 observation elements as defined in Table 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. Thus x_k and y_k belong to Poisson processes with the mean intensity of $\lambda_k A_k$ and $\pi_k \lambda_k B_k$. Where 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; 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). Formally,

$$X_\iota \sim Poisson(\pi_k \lambda_k B_k) \tag{2}$$

$$Y_\iota \sim Poisson(\lambda_k A_k) \tag{3}$$

where λ_k is an intensity that specifies the frequency of conflicts for element k. The conversion factor π_k, which specifies the probability that a conflict ends up with a crash in element k, can be obtained as follows:

$$\pi_k = \left(\frac{\sum x_{ki}}{\sum y_{ki}} \right) \frac{A_{ki}}{B_{ki}} \tag{4}$$

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 (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 3.2×10^{-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.

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 be increased, have no relation, or be decreased with the increase of observed conflicts. 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 (e.g. Williams, 1981; 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 π. Saunier and Sayed (2008) propose using an automated system to address issues that hinder the wide use of traffic conflict technique. Tarko and Songchitruksa (2005) proposed an extreme-value-theory approach to estimate the conversion factor (the π in Eq. (1)) based on measured post-encroachment time instead of crash and conflict counts. Davis et al. (2008) proposed a counterfactual approach to estimate the π in Eq. (1) by obtaining a probability an event could have been a crash, focusing on rear-end crashes. These more recent studies suggest a consensus emerging on the importance of π 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.

1.3. The desirable criteria for crash surrogate

While much of the historical crash surrogate research has been focused on traffic conflicts 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 and Songchitruksa, 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

Table 2
Final conversion factor (π) between conflicts and accidents (from Hydén (1987), Table 5.12).

	Car–car	Car-unprotected road user
Traffic class 1 + 2 (confidence interval)	3.2×10^{-5} (2.0×10^{-5}, 6.9×10^{-5})	15.3×10^{-5} (12.2×10^{-5}, 19.6×10^{-5})
Traffic class 3 + 4 (confidence interval)	11.1×10^{-5} (8.2×10^{-5}, 16.1×10^{-5})	3.2×10^{-5} (70.5×10^{-5}, 113.3×10^{-5})

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 attributes 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. *Surrogate 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 end point 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. Hydén, 1987; Chin and Quek, 1997); deceleration rate (e.g. Hydén, 1987); post-encroachment time (e.g. Hydén, 1987, 1996; Topp, 1998); deceleration-to-safety time (e.g. Topp, 1998); gap time, encroachment time, time-to-zebra (Várhelyi, 1996); proportion of stopping distance (FHWA, 2003); shock-wave frequency (VanArem and DeVos, 1997); "Jerks" (composite g-force and speed) (Gully et al., 1995); standard deviation of lateral position (Vogel, 2003); design consistency (IHSDM, 2008); time-line crossing (Vogel, 2003; Gordon et al., 2009); right-lane departure waning (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. (2010), and McGehee et al. (2010) are among the few who have recognized this explicit connection between context and surrogate metric.

1.4. Objective of this research

This paper 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 π, the crash-to-surrogate ratio, which forms the foundation for our approach. Data from a previous naturalistic driving study are used to assess our proposed structure. The paper concludes with an assessment of the implications of the research, particularly for analyses with naturalistic driving data.

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.

2.1. Analytical foundation of an event-based model approach

Fig. 1 is a conceptualization of the approach used to estimate π in Eq. (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 Eq. (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 π in Eq. (1).

The initial screening for naturalistic data has been traditionally conducted using vehicle kinematic triggers (see for example Dingus et al., 2005), but others 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 in order 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 Prob($Y_1 = 0$); i.e., an event that is *not* a crash or near crash of interest to us. 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 paper, the authors have a data set in which this classification has already occurred through video review of events.

The last step is the modeling of π. This should be conducted with a categorical outcome model such as a Logit or Probit (Amemiya, 1981; Agresti, 2002; Cameron and Trivedi, 2005).

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 π from Eq. (1), one

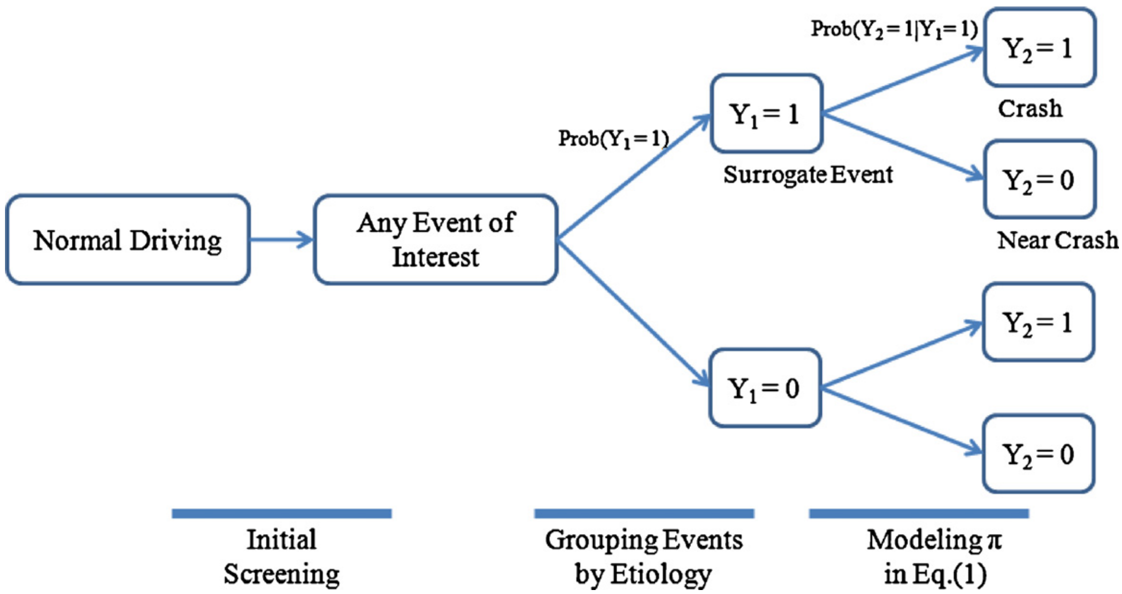

Fig. 1. Conceptualization of the relationship between crashes and near crashes in naturalistic driving data.

can quickly see the conceptual similarities in the two probabilities. The term π is the proportion of events of interest that are crashes (see Eq. (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 π 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 require 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 paper, not their application in safety management including exposure.

2.2. Estimating the conditional crash probability given a surrogate event

With respect to the conceptualization in Fig. 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, 2011; Wu and Jovanis, 2011). In our case, visual search of 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 Fig. 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:

$$P = \text{Prob}(Y_2 = 1 | Y_1 = 1, W, T) = \text{Prob}(CF_1 > CF_0) \tag{5}$$

$$CF_1 = V_1 + \varepsilon_1 = B_1 W + \Gamma T_1 + \varepsilon_1 \tag{6}$$

$$CF_0 = V_0 + \varepsilon_0 = B_0 W + \Gamma T_0 + \varepsilon_0 \tag{7}$$

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 ε 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 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 a near crash functions, respectively ($B_1 \neq B_0$). Γ vector, capture 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:

$$\begin{aligned} P &= \text{Prob}(Y_2 = 1 | Y_1 = 1, W, T) = \text{Prob}(CF_1 > CF_0) \\ &= \text{Prob}(\varepsilon_0 < \varepsilon_1 + V_1 - V_0) \end{aligned} \tag{8}$$

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

$$f(\varepsilon) = e^{-\varepsilon}e^{-e^{-\varepsilon}} \tag{9}$$

and the cumulative distribution is

$$F(\varepsilon) = e^{-e^{-\varepsilon}} \tag{10}$$

If ε_1 is given, then

$$P = \text{Prob}(\varepsilon_0 < \varepsilon_1 + V_1 - V_0) = e^{-e^{-(\varepsilon_1+V_1-V_0)}} \tag{11}$$

Since ε_1 is actually not given, the conditional crash probability is the integral of all P over all values of ε_1 weighted by its density function, $f(\varepsilon)$.

$$P = \int_{-\infty}^{\infty} (e^{-e^{-(\varepsilon_1+V_1-V_0)}})(e^{-\varepsilon_1}e^{-e^{-\varepsilon 1}})d\varepsilon_1 = \frac{e^{V_1}}{e^{V_1}+e^{V_0}} \tag{12}$$

Due to the identification issue for solving both B_1 and B_0, B_0 is normalized to zero, and hence:

$$CF_1 = B_1'W + \Gamma T_1 + \varepsilon_1 \tag{13}$$

$$CF_0 = \Gamma T_0 + \varepsilon_0 \tag{14}$$

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

$$P = \text{Prob}(Y_2 = 1|Y_1 = 1, W, T) = \frac{e^{V_1}}{e^{V_1}+e^{V0}} = \frac{e^{B_1'W+\Gamma T_1}}{e^{B_1'W+\Gamma T_1}+e^{\Gamma T_0}}$$

$$= \frac{e^{B_1'W+\Gamma(T_1-T_0)}}{e^{B_1'W+\Gamma(T_1-T_0)}+1} \tag{15}$$

And a logistic regression can be written as:

$$\log\left(\frac{P}{1-P}\right) = \log\left(\frac{e^{V_1}/(e^{V_1}+e^{V_0})}{1-e^{V_1}/(e^{V_1}+e^{V_0})}\right) = \log\left(\frac{e^{V_1}}{e^{V_0}}\right)$$

$$= V_1 - V_2 = B_1'W + \Gamma(T_1 - T_0) \tag{16}$$

An advantage of using this formulation of crash probability is that 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 Ω. The density of $\varepsilon_n' = (\varepsilon_0, \varepsilon_1)$ is

$$\phi(\varepsilon_n) = \frac{1}{(2\pi)^{1/2}|\Omega|^{1/2}}e^{-1/2\varepsilon_n'\Omega^{-1}\varepsilon_n} \tag{17}$$

The conditional crash probability is

$$P = \int I(\varepsilon_0 < \varepsilon_1 + V_1 - V_0)\phi(\varepsilon_n)\,d\varepsilon_n \tag{18}$$

where $I(.)$ is an indicator of whether the statement in parentheses holds, and the integral is over all value of ε_n. The integral does not have a closed form. For more detail, please refer to Train (2009).

2.3. Using conditional crash probability to estimate expected number of crashes

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

$$P_i = \text{Prob}(Y_2 = 1|Y_1 = 1, I = i) \tag{19}$$

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:

$$X_i = x_i \sim Binomial(N_i, P_i) \tag{20}$$

where x_i is the number of crashes observed for event scenario i, $x_i = 1, 2, \ldots, N_i$. P_i is the conditional crash probability given $Y_1 = 1$ and event scenario i. Therefore, given an event scenario i, if N_i surrogate events are observed, then the expected number of crashes is:

$$E(X_i) = N_iP_i \tag{21}$$

Considering P_i as the conditional crash probability given $Y_1 = 1$ and event scenario i, P_i can be written as:

$$P_i = \text{Prob}(Y_2 = 1|Y_1 = 1, I = i) = \frac{\text{Prob}(Y_2 = 1|Y_1 = 1, I = i)}{\text{Prob}(Y_2 = 1, I = i)}$$

$$= \frac{x_i/N}{N_i/N} = \frac{x_i}{N_i} = \pi_i \tag{22}$$

Therefore, P_i is a more generalized form of the crash-to-surrogate ratio, π, in Eq. (1). The whole derivation is summarized in Table 3. P_i represents the conditional crash probability for the surrogate events ($Y_1 = 1$) 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 Eq. (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 4, Table 7.

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 12–13 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.

Table 3
I independent binomials.

Event Scenario	Conditional crash probability	Surrogate Event Count (N)	Crash Count Generating Process	Expected Crash Count, $E(X)$
$i = 1$	P_1	N_1	$\text{Prob}(X_1 = x_1) = \binom{N_1}{x_1} P_1^{x_1}(1-P_1)^{N_1-x_1}$	$E(X_1)$
$i = 2$	P_2	N_2	$\text{Prob}(X_2 = x_2) = \binom{N_2}{x_2} P_2^{x_2}(1-P_2)^{N_2-x_2}$	$E(X_2)$
...	...	...	...	...
$i = I$	P_i	N_i	$\text{Prob}(X_i = x_i) = \binom{N_i}{x_i} P_i^{x_i}(1-P_2)^{N_i-x_i}$	$E(X_i)$

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 (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 4). Once the triggering events were found in the data, attributes of each event were saved from 30 s prior to the onset of the precipitating event, to 10 s after the completion of the event. Based upon the event criteria in Table 4, 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, 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.

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 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 5.

Table 4
Summary of kinematic search criteria for events in VTTI study (Dingus et al., 2005).

Trigger type	Description
1. Lateral acceleration	• Lateral accel. ≥0.7 g.
2. Longitudinal acceleration	• Accel. or decel. ≥0.6 g. • Accel. or decal. ≥0.5 and forward TTC ≤4 s • 0.4 g ≤longitudinal decel. <0.5 g, forward TTC ≤4 s, 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. ≥0.5 g and TTC ≤ 4 s • 0.4 g ≤longitudinal decel. <0.5 g, forward TTC ≤ 4 s, and forward range at the min. TTC ≤ 100 ft
5. Rear time-to-collision	• Rear TTC ≤ 2 s, rear range ≤50 feet, and absolute accel. of the following vehicle >0.3 g
6. Yaw rate	• Any value greater than or equal to a plus AND minus 4° change in heading (i.e., vehicle must return to the same general direction of travel) within a 3 s window of time

4. Application of method to road departure event analyses with naturalistic driving data

4.1. Model estimation

Fig. 2 highlights the portion of the conceptualized relationship between road departure crashes and near crashes (from Fig. 1) that is the focus of this paper. The conditional crash probabilities are estimated using a binary Logit model, since the alternative-specific

Table 5
Variable definitions.

Group	Variable	Definition	Variable type	Descriptive statistics
Dependent variable	Event outcome	Crash (1); near crash (0)	Binary	Crash = 13; near crash = 38
Situational characteristics	A straight trajectory before the event started	Go straight (1); otherwise (0)	Binary	Go straight = 32; otherwise = 19
	The type of trigger for the event is yaw rate, see Table 4	Yaw rate criteria was exceeded (1); otherwise (0)	Binary	Exceeded = 27; otherwise = 24
	The type of trigger for the event is lateral acceleration, see Table 4	Lateral acceleration criteria was exceeded (1); otherwise (0)	Binary	Exceeded = 11; otherwise = 40
	Maximum absolute value of lateral acceleration before the end of an event	Continuous variable (g-force)	Continuous	Max = 2.2 g; min = 0.06 g
	Event duration	Continuous variable (s)	Continuous	Max = 12.4 sec; min = 1.8 sec
	Road departure (left) before the event	Road departure (1); otherwise (0)	Binary	Departure = 13; otherwise = 38
	Road departure (right) before the event	Road departure (1); otherwise (0)	Binary	Departure = 28 otherwise = 23
	Driver was distracted	Distracted (1); otherwise (0)	Binary	Distracted = 39; otherwise = 12
	The presence of driver fatigue	Driver fatigue(1); otherwise (0)	Binary	Fatigue = 18; otherwise = 33
Driving context	Event occurred on a horizontal curve	Curve (1); otherwise (0)	Binary	Curve = 21; otherwise = 30
	The presence of daylight	Daylight (1); otherwise (0)	Binary	Daylight = 29; otherwise = 22
	Dry pavement surface	Dry (1); wet/icy/snowy (0)	Binary	Dry = 40; wet/snowy = 11
	Roadway with median	Divided (1); otherwise (0)	Binary	Divided = 23; otherwise = 28
	The event was occurred in a rural area	Rural area (1); otherwise (0)	Binary	Rural area = 24; otherwise = 27

Table 6
The estimated event-based models.

Group	Variables	Model 1		Model 2		Model 3	
		Coef.	Odds ratio	Coef.	Odds ratio	Coef.	Odds ratio
	The type of trigger for the event is yaw rate	−1.341	0.262				
	S.E.	1.539	0.403	NA		NA	
	p-value	0.384					
	Lateral acceleration rate greater than 0.7 g	2.543	12.723	3.193	24.353	1.216	3.374
	S.E.	1.582	20.132	1.449	35.284	0.951	3.208
	p-value	0.108		0.028		0.201	
	Event duration	−0.021	0.979	NA		NA	
	S.E.	0.028	0.027				
	p-value	0.441					
	A straight trajectory before the event started	−1.775	0.169	−1.882	0.152	−0.629	0.533
	S.E.	1.233	0.209	1.16	0.177	0.801	0.427
Situational characteristics	*p*-value	0.15		0.105		0.433	
	Road departure (left) before the event	0.913	2.492	NA		NA	
	S.E.	2.196	5.472				
	p-value	0.677					
	Road departure (right) before the event	0.453	1.573	NA		NA	
	S.E.	1.64	2.581				
	p-value	0.782					
	Driver was distracted	−0.071	0.932	NA		NA	
	S.E.	1.375	1.281				
	p-value	0.959					
	The presence of driver fatigue	0.694	2.002	NA		NA	
	S.E.	1.06	2.123				
	p-value	0.513					
	Dry pavement surface	−2.739	0.065	−2.258	0.105	−1.344	0.261
	S.E.	1.365	0.088	1.091	0.114	0.794	0.207
	p-value	0.045		0.039		0.091	
	Roadway with median	−1.832	0.16	−1.214	0.297	−0.848	0.428
	S.E.	1.514	0.242	0.961	0.285	0.852	0.365
	p-value	0.226		0.206		0.32	
	Presence of a horizontal curve	−3.166	0.042	−3.23	0.04	NA	
	S.E.	1.488	0.063	1.424	0.056		
Driving context	*p*-value	0.033		0.023			
	The event was occurred in a rural area	−1.742	0.175	−1.778	0.169	−1.171	0.31
	S.E.	1.008	0.177	0.943	0.159	0.795	0.246
	p-value	0.084		0.059		0.141	
	The presence of daylight	−1.528	0.217	−1.764	0.171	−0.851	0.427
	S.E.	1.227	0.266	1.008	0.173	0.809	0.346
	p-value	0.213		0.08		0.293	
	Constant term	5.822	NA	4.317	NA	1.227	NA
	S.E.	2.62		1.929		0.961	
	p-value	0.026		0.025		0.202	
	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 ration test (chi-square)	21.21		19.22		10.86	
	Likelihood ration test (*p*-value)	0.069		0.0075		0.0929	

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, allows us to estimate π, 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.

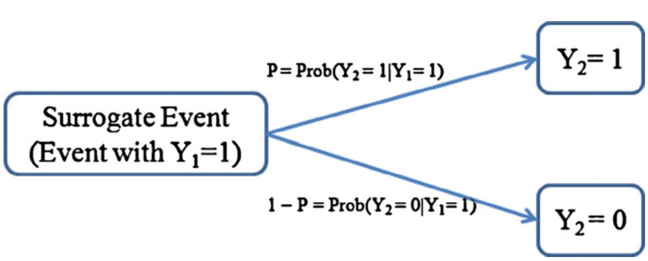

Fig. 2. Conditional crash probability for a surrogate event.

After the search of potential covariates described in Section 3 a model was estimated using significant predictors from all the testing; this model is shown as model 1 in Table 6. The remaining 2 models show the results of removal of 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.

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

Table 7
Conditional crash probabilities based on VTTI 100-car single vehicle run-off-road events.

1 Context	2 Go straight	3 Lateral accel. trigger	4 Dry pave.	5 Divide road	6 Horiz. curve	7 Rural area	8 Day light	9 Cond. crash prob.	10 Observed surrogate events	11 Expected crashes	12 Observed crashes
1	1	0	1	1	1	1	1	0.0004	1	0.000	0
2	1	0	1	1	1	0	1	0.0024	1	0.002	0
3	0	0	1	1	1	1	1	0.0027	1	0.003	0
4	0	0	1	0	1	1	1	0.0089	1	0.009	0
5	0	0	1	1	1	1	0	0.0153	1	0.015	0
6	1	0	1	0	0	1	1	0.0334	3	0.100	0
7	1	0	1	0	1	0	0	0.0451	3	0.135	1
8	0	0	1	0	1	0	1	0.0504	1	0.050	0
9	1	1	1	1	1	1	0	0.0545	1	0.055	0
10	1	0	1	1	0	1	0	0.0565	1	0.056	0
11	1	0	1	1	0	0	1	0.0572	5	0.286	0
12	0	0	1	1	0	1	1	0.0631	1	0.063	0
13	0	0	0	1	1	1	0	0.1294	1	0.129	0
14	1	0	1	0	0	1	0	0.1678	3	0.503	0
15	1	0	1	0	0	0	1	0.1698	1	0.170	0
16	0	1	1	0	1	1	1	0.1794	1	0.179	0
17	0	0	1	0	0	1	1	0.1849	2	0.370	1
18	1	1	1	1	0	1	1	0.1999	3	0.600	1
19	0	0	1	0	1	0	0	0.2366	1	0.237	0
20	1	0	1	1	0	0	0	0.2616	2	0.523	1
21	0	1	1	1	1	1	0	0.2746	1	0.275	0
22	0	0	1	1	0	0	1	0.2850	1	0.285	1
23	0	0	0	0	1	1	0	0.3337	1	0.334	0
24	0	0	0	0	1	0	1	0.3370	2	0.674	1
25	1	1	0	1	1	0	1	0.3588	2	0.718	0
26	1	1	1	0	1	0	0	0.5348	1	0.535	1
27	1	0	1	0	0	0	0	0.5441	2	1.088	0
28	0	0	1	0	0	0	1	0.5732	1	0.573	0
29	1	1	0	0	1	1	0	0.6501	1	0.650	1
30	1	0	0	0	0	1	0	0.6585	1	0.659	1
31	0	0	1	0	0	0	0	0.8868	1	0.887	1
32	1	0	0	0	0	0	0	0.9195	1	0.919	1
33	0	0	0	0	0	0	1	0.9278	1	0.928	1
34	0	1	0	1	0	0	1	0.9893	1	0.989	1

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 0.7 g 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 1.8–12.4 s duration of the event itself; during this short period of time, the event probability increases by this 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 6. This finding supports the intuition that dry surface can reduces vehicle stopping distance, and hence reduces 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 types of areas respectively. Examining of the data set, found that there are only 4 crashes but 17 near crashes that occurred with the presence of a horizontal curve; hence the estimated odds ratio is close to zero (see odds ratio for the presence of a horizontal curve in Table 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.

4.2. Estimating the expected number of crashes using surrogate events

Table 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 2–8 of Table 7. The estimated conditional crash probabilities from Model 3 (appearing in column 9) and observed surrogate events for the specific context from the data (column 10) are applied using the binomial models of Table 3 and Eq. (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 2–8 of Table 7: the vehicle's trajectory is not straight; the vehicle's maximum lateral acceleration is less than 0.7 g, 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. Actually, one of these two surrogate events ended up as 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 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 31–34 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 US, 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.

5. Summary and discussion

This paper seeks to provide a framework that will facilitate the use of surrogates in road safety analyses. The paper 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 π, the crash to surrogate ratio (see Eq. (1)). A conceptual structure for the estimation of π 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 (e.g. meeting the screening criteria listed in Table 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 0.7 g 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 7).

While the results are promising, there are limitations to the study. The study 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.

References

Agresti, A., 2002. Categorical Data Analysis, 2nd edition. Wiley.

Amemiya, T., 1981. Qualitative response models: a survey. Journal of Economic Literature 19, 1483–1536.

Amundsen, F.H., Hyden, C., 1977. Proceeding of First Workshop on Traffic Conflicts. Institution of Transport Economics, Oslo/Lund Institute of Technology, Oslo, Norway.

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.

Battelle, 2006. Evaluation of the Volvo Intelligent Vehicle Initiative Field Operational Test (No. version 1.3). Washington, DC.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

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

Cooper, P., 1984. Experience with traffic conflicts in canada with emphasis on post encroachment time techniques. In: Asmussen, E. (Ed.), International Study of Traffic Conflict Techniques. Springer-Verlag, Berlin, pp. 75–96.

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

Davis, G.A., Hourdos, J., Xiong, H., 2008. Outline of a causal theory of traffic conflicts and collisions. In: Proceedings of the 87th Transportation Research Board Annual Meeting (CD-ROM), National Research Council, TRB, Washington, DC.

Davis, G.A., Swenson, T., 2006. Collective responsibility for freeway rear-ending accidents: an application of probabilistic causal models. Accident Analysis and Prevention 38 (4), 728–736.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R., 2005. The 100-Car Naturalistic Driving Study, Phase II—Results of the 100-Car Field Experiment. National Highway Traffic Safety Admin. (Contract No. DTNH22-00-C-07007).

FHWA, 2003. Surrogate Safety Measures from Traffic Simulation Models. Final Report, Publication No. FHWA-RD-03-050, Federal Highway Administration.

FHWA, 2008. Interactive Highway Safety Design Model (IHSDM).

Fitch, G.M., Rakha, H.A., Arafeh, M., Blanco, M., Gupta, S.A., Zimmermann, R.P., Hanowski, R.J., 2008. Safety Benefit Evaluation of a Forward Collision Warning System. US Department of Transportation, DOT HS 810 910. Washington, DC.

Gordon, T., Green, P., Kostyniuk, L., Barnes, M., Blankespoor, A., Bogard, S., Blower, D., 2009. A multivariate analysis of crash and naturalistic event data in relation to highway factors using the GIS framework. In: Proceedings of the 4th SHRP2 Safety Research Symposium, Washington, DC.

Grayson, G.B., Hakkert, A.S., 1987. Accident analysis and conflict behaviour. In: Rothengatter, J., de Bruine, R.A. (Eds.), Road-user and Traffic Safety. Van Gorcum, pp. 27–59.

Gully, S.M., Whitney, D.J., Vanosdall, F.E., 1995. Prediction of police officers traffic accident involvement using behavioral observations. Accident Analysis and Prevention 27 (3), 355–362.

Greene, W.H., 2003. Econometric Analysis, 5th 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 2147, 66–74.

Hauer, E., 1982. Traffic conflicts and exposure. Accident Analysis and Prevention 14 (5), 359–364.

Hauer, E., Gårder, P., 1986. Research into the validity of the traffic conflicts technique. Accident Analysis and Prevention 18 (6), 471–481.

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.

Hydén, C., 1996. Traffic conflicts technique: state-of-the-art. In: Topp, H.H. (Ed.), Traffic Safety Work with Video-processing. University Kaiserslautern. Transportation Department, Green Series No. 43.

Jovanis, P.P., Agüero, K., Wu, F., Shankar, V., (2011). Naturalistic driving event data analysis: omitted variable bias and multilevel modeling approaches. Journal of the Transportation Research Board, in press.

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.

McFadden, D., 1974. Conditional Logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

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.

Migletz, D., Glauz, W., Bauer, K., 1985. Relationships between Traffic Conflicts and Accidents, Report FHWA/RD-84/042. Federal Highway Administration, Washington, DC.

Oppe, S., 1986. Evaluation of traffic conflict techniques. In: Proceeding of the Workshop on Traffic Conflicts and Other Intermediate Measures in Safety Evaluation, Budapest.

Perkins, S.R., Harris, J.L., 1967. Traffic Conflict Characteristics: Accident Potential at Intersections. General Motors Corporation, Warren, MI.

Risser, R., 1985. Behavior in traffic conflict situations. Accident Analysis and Prevention 17 (2), 179–197.

Saunier, N., Sayed, T., 2008. A probabilistic framework for the automated analysis of the exposure to road collision. Transportation Research Record 2019, 96–104.

Sayed, T., Zein, S., 1999. Traffic conflict standards for intersections. Transportation Planning and Technology 22, 309–323.

Shankar, V., Jovanis, P.P., Aguero, J., Gross, F., 2008. Analysis of naturalistic driving data: a prospective view on methodological paradigms. Transportation Research Record 2061, 1–8.

Strategic Highway Research Program 2, 2010. Website for Naturalistic Driving Data: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Pages/RFP_S08_Resources_and_Reference_Material_487.aspx.

Svensson, Å., 1998. A Method for Analyzing the Traffic Process in a Safety Perspective. Bulletin 166. Dept. of Traffic Planning and Engineering, Lund University, Lund, Sweden.

Tarko, A.P., Songchitruksa, P., 2005. Estimating frequency of crashes as extreme traffic events. In: Proceedings of the 84th Transportation Research Board Annual Meeting (CD-ROM), National Research Council, TRB, Washington, DC.

Tarko, A.P., Davis, G, Saunier, N., Sayed, T., Washington, S., 2009. Surrogate Measures of Safety, White Paper.

Topp, H.H., 1998. Traffic Safety Work with Video-Processing. University Kaiserslautern, Transportation Department, Green Series No. 43, Kaiserslautern, Germany.

Train, K., 2009. Descrete Choice Methods with Simulation, 2nd edition. Cambridge University Press.

Várhelyi, A., 1996. Dynamic Speed Adaptation Based on Information Technology—A Theoretical Background, Bulletin 142. Dept. of Traffic Planning and Engineering, Lund University, Lund, Sweden.

VanArem, B., DeVos, A.P., 1997. The Effect of a Special Lane for Intelligent Vehicles on Traffic Flows. TNO-INRO Report 1997-02a. Delft, The Netherlands.

Vogel, K., 2003. Modeling Driver Behavior–A Control Theory Based Approach. Institute of Technology, University of Linköping, Linköping Studies in Science and Technology, Dissertation No. 751. Sweden.

Volvo, 2005. Volvo Trucks Field Operational Test: Evaluation of Advanced Safety Systems for Heavy Truck Tractors. Washington, DC.

Williams, M.J., 1981. Validity of the traffic conflicts technique. Accident Analysis and Prevention 13 (2), 133–135.

Wu, K., Jovanis, P.P., 2011. Defining, screening, and validating crash surrogate events using naturalistic driving data. In: Proceedings of the 3rd International Conference on Road Safety and Simulation.