Check for updates

Analysis of Naturalistic Driving Event Data

Omitted-Variable Bias and Multilevel Modeling Approaches

Paul P. Jovanis, Jonathan Aguero-Valverde, Kun-Feng Wu,
and Venky Shankar

Naturalistic driving studies have been conducted over the past 5 years or more and have commonly reviewed video and kinematic data to identify and analyze crash, near-crash, and critical-incident events. But statistical methods that are applicable to these event data are needed. This paper addresses two issues in model development for naturalistic driving event data: the test for omitted-variable bias and the exploration of the advantages of hierarchical model structures in data analysis. With roadway departure event data from the 100-Car Naturalistic Driving Study conducted at Virginia Tech Transportation Institute, Blacksburg, Virginia, logit models were used to estimate the probability that a crash or a near crash would occur, rather than a critical incident. The models indicated a substantial omitted-variable bias for estimation of the effect of context variables but little difference for driver variables. These tests indicated that modeling of naturalistic event data should have included variables that described the attributes of the event, the driver, and the context to reduce the likelihood of bias. Hierarchical model structures offer the advantage of driver-level predictors to parameterize the effects of event attributes and contexts. The models thus reflect how driver decisions are executed: drivers with particular characteristics (one level) find themselves in contexts in which they execute specific driving maneuvers (second level), which lead to certain outcomes. Suggestions for further research include testing with additional data sets and potential applications to analysis of crash surrogates.

For many years, road safety studies have been conducted by using data collected at the scene after a crash event has occurred. Law enforcement officers measure and photograph the scene and collect data from involved participants (e.g., drivers, witnesses, pedestrians) who may have substantial self-interest. Typically, it is difficult to understand the sequence of actions that preceded the crash, because involved parties may not remember events accurately and witnesses may be unreliable (*1*). The crash data (including environmental characteristics at the time of the crash) developed through this process are then combined with roadway inventory data, which concern traffic volume and roadway characteristics, to build statistical models of the probability of having 0, 1, 2, . . . crashes on a road segment within a period of time (*2–5*).

P. P. Jovanis, K.-F. Wu, and V. Shankar, Department of Civil and Environmental Engineering, Larson Institute, Pennsylvania State University, 212 Sackett Building, University Park, PA 16802-1408. J. Aguero-Valverde, Programa de Investigación en Desarrollo Urbano Sostenible, Universidad de Costa Rica, Barrio Los Profesores, Calle B, No. 11, Mercedes San Pedro, San José 11503, Costa Rica. Corresponding author: P. P. Jovanis, Ppj2@engr.psu.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2236, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 49–57.
DOI: 10.3141/2236-06

In response to the uncertainties that surround crash etiology, naturalistic driving has been applied in a number of studies. There are two primary distinguishing features of this method:

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 onboard sensors continuously record vehicle accelerations in three dimensions as well as rotational motion along the same axes. Radars are often present to record proximity to other vehicles and potential obstacles on the roadway or roadside.
2. Drivers are asked to drive as they normally would (i.e., without specific experimental or operational protocols and not in a simulator or test track). The period of observation can vary from several weeks to a year or more.

All these data are recorded and stored within an onboard data acquisition system (DAS). The DAS for each vehicle is periodically copied into a searchable database and assembled for later analysis. Rather than rely on law enforcement officer judgment or witness recollection, the DAS can be used to record virtually all actions of the subject driver before, during, and after each event. Because events are recorded by using video and vehicle sensors, individual events of interest can generally be described with greater accuracy and reliability than they can be by crash reports assembled after the fact.

One common way to analyze naturalistic data is to identify events, in addition to crashes, that represent a range of near-crash conditions. It is hoped that these additional events can be used to supplement crash observations and generally improve safety analyses (e.g., as surrogates for crashes). Typically, crashes and near crashes are identified through the detection of unusual vehicle kinematics recorded electronically through accelerometers and gyroscopic sensors. Table 1 shows an example of search criteria used to identify events for the 100-Car Naturalistic Driving Study conducted at Virginia Tech Transportation Institute (VTTI), Blacksburg, Virginia. (*6*). Vehicle-based accelerometers are used to measure lateral and longitudinal acceleration; these measures are used individually or with time-to-collision 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. Yaw rate level is used to identify large heading changes within a short time period, which are another potential event indicator. Last, forward and rear time to collision can be used with vehicle kinematics (including measurements of a target vehicle) to identify additional events. Once identified kinematically, the events are screened through use of forward and face video. If verified as safety-related events, they are retained. If not, they are discarded. Data for the period shortly before, during, and shortly after the event are preserved. The result is

TABLE 1 Summary of Kinematic Search Criteria for Events in VTTI Study (*6*)

Trigger Type	Description
Lateral acceleration	Lateral acceleration ≥ 0.7 g.
Longitudinal acceleration	Acceleration or deceleration ≥ 0.6 g. Acceleration or deceleration ≥ 0.5 and forward TTC ≤ 4 s. 0.4 g ≤ longitudinal deceleration < 0.5 g, forward TTC ≤ 4 s, and forward range at minimum TTC ≤ 100 ft.
Event button	Activated by driver by pressing button on dashboard when event occurred that he or she deemed critical.
Forward TTC	Acceleration or deceleration ≥ 0.5 g and TTC ≤ 4 s. 0.4 g ≤ longitudinal deceleration < 0.5 g, forward TTC ≤ 4 s, and forward range at minimum TTC ≤ 100 ft.
Rear TTC	Rear TTC ≤ 2 s, rear range ≤ 50 ft, and absolute acceleration of the following vehicle > 0.3 g.
Yaw rate	Heading change ≥ 4° within 3-s window (vehicle must return to same general direction of travel).

NOTE: TTC = time-to-collision.

a set of potentially rich data that offers insights into crashes and near crashes that previously were unavailable. Although 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 time, including in crash and near-crash involvements. It is the analysis of these events that is central to this paper.

Naturalistic driving has been applied to studies of drivers from the regular driving population (*6, 7*), truck drivers (*8–10*), young drivers, and older drivers (*11*). A series of technology tests of onboard safety equipment has used the naturalistic technique also (*12–14*).

Reviews of video in and around a vehicle can be used to enhance the electronically based measures with variables, which typically are recorded on police accident reports (e.g., crash type, assessment of precipitating event, driver distraction, impairment). In addition, driver demographic and physiologic information is collected at project initiation along with some type of driving style attributes, which often are measures of possible crash predisposition, such as indices of driver aggression or life stress. Table 2 contains a summary of these data types, including variable types, description, and source of information. The table is intended to provide an introduction to the types of variables typically collected during a naturalistic driving study.

TABLE 2 Typical Variables That Describe Naturalistic Driving Events

Variable Type	Description	Source	Comment
Dependent variable: event of interest	Crash; near crash; critical incident; noncrash	Observed from video (see text)	Crashes include events recorded on police accident reports; others are new information available only from naturalistic studies.
Event attributes			
Precipitating event	Event immediately preceding crash	Video	Observed from video in naturalistic studies; derived by law enforcement at scene by using judgment and crash investigation and reconstruction.
Driver impairment	Distraction; alcohol or drug involvement	Video	Distraction is observed in naturalistic studies; derived with great difficulty by law enforcement at scene of crash; alcohol or drug involvement observed in naturalistic study but with much judgment (unless alcohol sensor used). Measured or tested in field by law enforcement for crash events only.
Driving context: road, environment, and traffic conditions at time of event	Presence of road element or environmental condition at time of event	Video, GIS	In naturalistic studies the context within which the event occurred may be observed through the use of video in combination with GIS (if available). May also be verified through other sensor measures.
Driver attributes			
Demographic (e.g., gender, years driving, age)	Self-reported demographic data (e.g., age, years driving)	Self-reported survey	Obtained through self-reports as recorded on questionnaires before initiation of driving in instrumented vehicle.
Physiological (e.g., visual or other impairments)	Assessment of various physiological conditions at time of initiation into study	Self-reported or test result during study enrollment	Obtained through self-reports during study enrollment.
Psychological (measures of crash predisposition)	Measures of personality, life stress, or risk acceptance obtained during study enrollment	Reported through use of specific tools before driving	Specific predisposition measures may be used; most involve response to specific set of questions aimed at assessing the extent of predisposition (e.g., measures of driving aggression).

NOTE: GIS = geographic information system.

STUDY GOALS

This paper builds on earlier research (*15*) and seeks to provide a framework for the analysis of the event-based naturalistic driving data. Although substantial progress has been made in developing techniques to identify events of interest and describe the events qualitatively, more limited progress has been made in developing and testing analysis paradigms that apply to common naturalistic data (*16*) (e.g., the development and testing of models that can be systematically applied to data associated with the identified events of interest to gain insight into crash etiology and contributing factors).

One particular issue that concerns naturalistic event data modeling is the effect of omitted-variable bias. The bias occurs when a relevant variable is omitted from a specified model (*17*). Given the limited analysis of naturalistic event data, this seems a logical initial step.

An additional issue is the potential application of Bayesian hierarchical models. Shankar et al. point out that hierarchy can be used to reflect how driver decisions are made: driver-level predictors (e.g., years of driving) can be used to parameterize the effects of event attributes and context (*15*). The structure of the models reflects how driver decisions are executed: drivers with particular characteristics (one level) find themselves in contexts in which they execute specific driving maneuvers (second level), which lead to certain outcomes. The next section of the paper describes the methodology used to study the naturalistic events. A description follows of the particular data used in this analysis, namely, the VTTI 100-car data set. The results are discussed, and conclusions and suggestions for future research end the paper.

METHODOLOGY

Analysis Approach

Figure 1 summarizes the hypothesized relationship between the outcomes (i.e., crashes, near crashes, and critical incidents) and the three components conceptualized as contributing to the outcomes (i.e., context, driver attributes, and event attributes). With respect to outcomes,

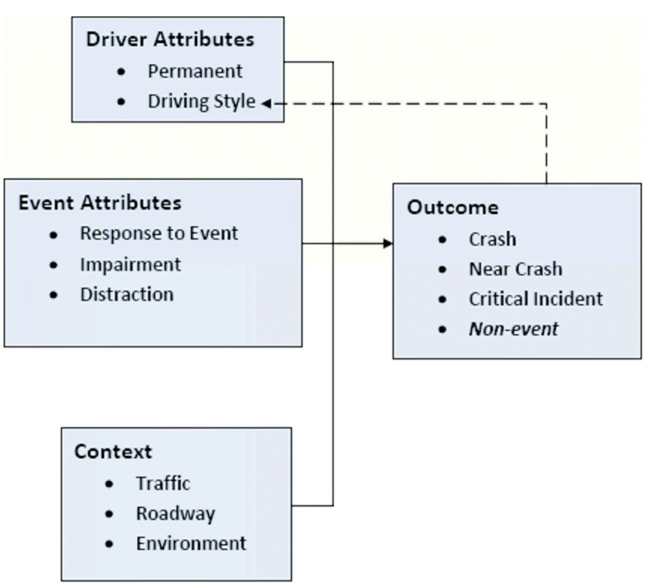

FIGURE 1 Study analysis approach.

a crash is defined as "any measurable dissipation or transfer of energy due to the contact of the subject vehicle with another vehicle or object" (*6*). A near crash is an event in which a crash is avoided only through an action that approaches the operating limits of the driver or vehicle by means of a rapid maneuver. "As a guide: subject vehicle braking greater than 0.5 g, or steering input that results in a lateral acceleration greater than 0.4 g to avoid a crash, constitutes a rapid maneuver" (*6*). A critical incident is one in which a crash is avoided by an extreme steering or braking input (or both) that does not approach driver or vehicle limits. This classification is conducted after use of initial event screening criteria, such as those shown in Table 1. An outcome of nonevents has been added to Figure 1; nonevents are periods of driving in which none of the Table 1 event criteria applies. Nonevents can be thought of as normal or baseline driving. This outcome is shown in italics in the figure because it was given to the research team without predictor variables. Thus it could not be used in the paper. The conclusions and recommendations offer some advice about modeling that can be conducted with such outcomes.

Context includes descriptors of the roadway, traffic level, and environmental conditions at the time of the event. Driver attributes include permanent characteristics, such as age and years of driving experience (constant over the course of the year or so of a naturalistic study), and driving style, which is intended to convey the level of risk the driver is willing to accept while undertaking the driving task. Event attributes include descriptors of the situation as it unfolded during the event (such as driver response recorded by onboard sensors). Driver attributes, such as impairment and distraction, captured in the few seconds around the crash are included as event attributes. Typically, they are retained for analysis in the seconds before and during an event; no observation is implied for driver impairment or distraction at other times. One way to retain the event-based method and add observations beyond the immediate event time is to randomly collect nonevent driver observations throughout the driving time (*6*). This would effectively add a fourth outcome to the conceptual framework, that is, a nonevent outcome for a driving period. As these nonevent driving periods were not available to the research team at the same level of detail as the other three outcomes, no modeling could be conducted. Extensions to the modeling, which include these outcomes, are included in the discussion of future research needs.

Each of these four components is depicted as contributing to the event outcome. In addition, the dashed line in Figure 1 attempts to illustrate that a driver's experience with crash outcomes may, after some period of time, affect his or her driving style. It is not expected that this will occur during the typical duration of a naturalistic study (1 year or less) but for longer-term studies it is possible that such a hypothesized change may be observable in the data set. The dashed line is included here to recognize that such driver adaptation may occur, but it is not explicitly modeled in this study.

Events Analyzed as Categorical Outcomes

Consistent with the fact that the events may be viewed as discrete outcomes, a range of categorical model forms may be used to compare crash, near-crash, and critical incident outcomes, including binary logit, multinomial logit, and ordered logit (*16*). A comparison of each model formulation yielded generally consistent parameter signs, magnitudes, and levels of statistical significance. Only the binary logit results are reported, because they are the most easily interpreted (*16*). A direct extension of this formulation is to compare any pair of outcomes (e.g., crashes as opposed to nonevents without any extreme maneuvers), if sample sizes support such formulations. The analyses in this study were constrained by sample size limitations (see next section) so a conditional formulation was developed: one outcome consists of crashes and near crashes; the other outcome is the set of critical incidents.

In the binary logit formulation, there are only two possible outcomes, so the base alternative was chosen to be a critical incident, and the other alternative was a crash or near crash. Positive parameter signs in this model reflect increases in the likelihood of an event being a crash or near crash instead of a critical incident. This model can be described by the following equation:

$$\mathrm{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_{k=1}^{K} \beta_k X_k \tag{1}$$

where

- p_i = probability of success for event i (crash or near crash is considered a success),
- α = intercept,
- K = highest value of covariate k,
- β_k = coefficient for covariate k, and
- X_{ik} = covariate k for event i.

Goodness-of-fit statistics for the logit model included the log likelihood and pseudo-R^2 statistic. The binary logit proved to have the best overall fit and the most interpretable results. Thus it was chosen for discussion in this paper.

Such a formulation identifies the factors that differentiate more severe outcomes (crashes and near crashes) from those that are less severe (critical incidents). The testing allowed an exploration of how the three components in Figure 1 interacted to lead to the severe outcomes, as compared with those that were less severe. There was a rough analogy to a comparison of crash severity by using police accident reports. The difference lay in lack of experience in formulating naturalistic, event-based models; it is hoped that this research will contribute to the body of knowledge that needs to be built concerning such analyses.

To develop a clearer understanding of how each predictor type influenced the dependent variable, each set of predictors from the three factors shown in Figure 1 was tested individually. Once the individual factors were modeled, pairs were entered in the model, and finally a model with all four sets of predictors was entered at one time. This series of models allowed the test of any changes in predictor significance magnitude and sign. Given the limited number of naturalistic driving analysis studies that have been undertaken, this cautious, building block approach seemed advisable. The remainder of this section describes the way in which the events were modeled as outcomes.

Models were estimated with just one type of variable entered at a time. This resulted in three models: one with driver attributes only, one with event attributes only, and one with context variables only. Additional models were estimated with two of the three of the components included. A last model included all three types of variables. This method allowed for careful evaluation of the effect of omitted-variable bias. The team could track changes in estimated parameters as the series of models was developed.

Multilevel Formulation

One weakness of the standard binary logit model is that it treats each observation of outcome as independent. This ignores the fact that individual drivers could have been involved in multiple events of the same type, and it does not allow the effects of crash contributing factors to vary across drivers. The advantage of using a model hierarchy that reflects driver decisions and outcomes is the additional advantage of this approach. The model depicted in Figure 2 specifically recognizes events as occurring at the first level but then associates those events with specific driver attributes at the second level. This model can be formulated as a two-level, random intercept one with the driver at the higher level. If the intercept is allowed to vary randomly across drivers,

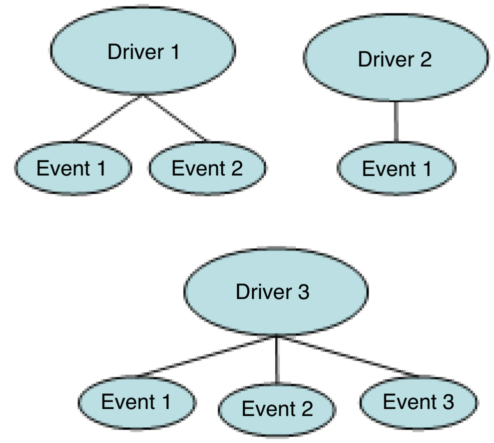

FIGURE 2 Hierarchical structure for event-based model.

$$\mathrm{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_{0j} + \sum_{k=1}^{K} \beta_k X_k \tag{2}$$

$$\alpha_{0j} = \alpha_0 + u_{0j} \tag{3}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \tag{4}$$

where

- p_{ij} = probability of success for event i of driver j (where a crash or near crash is considered a success),
- α_{0j} = intercepts (one for each driver),
- β_k = coefficient for covariate k, and
- X_{ik} = covariate k for event i.

In a two-level, random intercept model, the intercept consists of two terms: a fixed component α_0 and a driver-specific component, the random effect u_{0j}, which is assumed to follow a normal distribution with mean zero and standard deviation σ_{u_0}.

Additional goodness-of-fit statistics may be used with multilevel formulations. The variance partition coefficient, which comes from a two-level random intercept model, is the proportion of total residual variance that is attributable to Level 2. In a binary logit model, an alternative variance partition coefficient measure is needed, because the variance partition coefficient itself will depend on the explanatory variable (*18*). Snijders and Bosker (*19*) suggest that, for a binary logit model, the variance partition coefficient can be computed as

$$\frac{\sigma_{u_0}}{\sigma_{u_0} + \frac{\pi^2}{3}} = \frac{\sigma_{u_0}}{\sigma_{u_0} + 3.29} \tag{5}$$

All models were implemented by using MLwiN 2.0 software (*18*). Maximum likelihood estimation in a binary logit multilevel model is computationally intensive, so the quasi-likelihood method is implemented in MLwiN. This procedure uses a linearization method, which incorporates a Taylor series expansion to transform a binary response into a continuous response. The model is then estimated by using iterative, generalized least squares, or reweighted iterative, generalized least squares (*18*).

The receiver operating characteristic curve was used to compare the final two models. The receiver operating characteristic is a standard technique that compares the true positive rate for a classification with the false positive rate (*20*). The area under the receiver operating characteristic is also referred to as the index of accuracy or concordance index. Larger areas under the curve indicate better model prediction power (*21*).

DATA

The VTTI 100-Car Naturalistic Driving Study data set was used for this study (*6*). It includes 241 primary and secondary drivers, and 12 to 13 months of data collection for each vehicle. A DAS was installed in each vehicle and consisted of cameras for video recording, kinematic sensors, radar, lane tracking devices, and a hard drive for data storage (*22*). At certain points during the study, information from the DAS hard drive was received by VTTI, and triggering software was run on the basis of event criteria (see Table 1). Once certain triggers were found in the data, attributes of each event were saved from 30 s beforehand to 10 s after the onset of the precipitating event.

On the basis of the event criteria in Table 1, VTTI researchers identified 69 crashes, 761 near crashes, and 8,295 critical events over the entire study. These events of interest, as a whole, spanned the full range of crashes and related events, including intersection crashes and roadway departure events. Because the focus of this study was road departure events, the sample size was reduced to 22 crashes, 36 near crashes, and 170 critical incidents.

Precipitating factors occurred outside the vehicle only and did not include driver distractions or impairments (consistent with having a structure similar to a police accident report). Because the only events analyzed in the data set were run-off-road events, the only precipitating factors included were those that involved vehicle loss of control and crossover of lane lines or roadway edges. Cameras installed inside each vehicle were used to observe driver facial features and head movements, which allowed for the recording of certain distractions or impairments. Driver impairments were checked by using a few methods, which depended on the type of impairment of concern. By using video, it was possible to directly observe impairment related to drowsiness, fatigue, sleepiness, or the state of being asleep, as well as some cases of drug or alcohol use (when it was explicit). In other cases of drug or alcohol use, however, the driver had to admit to the behavior after the study ended. Distractions were noted if drivers undertook any secondary tasks during the time period before a precipitating factor occurred. Twelve primary distraction categories were included in the data set, with a total of 40 that were on the basis of specific distractions that belonged to a given group. These distractions were grouped into seven final categories (Table 3).

TABLE 3 Variable Definitions

Variable	Definition	Variable Type	Mean	SD	Min.	Max.
Dependent variable: event outcome	Crash or near crash (1); critical incident (0)	Binary	0.251	0.435	0	1
Event attributes						
Precipitating Factor 1. Loss of control	Lose control of vehicle because of vehicle failures, poor road conditions, excessive speed (from GES critical event)	Binary	0.370	0.233	0	1
Precipitating Factor 2. Subject over lane line or road edge	System detected the vehicle over the lane line or roadway edge (from GES critical event)	Binary	0.498	0.250	0	1
Driver Impairment 1. Drowsy, sleepy, asleep, fatigue	Driver appears to show these characteristics (all are based on GES "driver distracted by" variable)	Binary	0.207	0.164	0	1
Distraction 1. Wireless device	Distraction related to locating or operating a wireless device	Binary	0.093	0.084	0	1
Distraction 2. Vehicle related	Adjusting climate control, radio, audio devices, etc.	Binary	0.044	0.042	0	1
Distraction 3. Passenger related	Distraction attributable to passenger in vehicle	Binary	0.057	0.054	0	1
Distraction 4. Talking, singing, or daydreaming	Self-evident definition	Binary	0.035	0.034	0	1
Distraction 5. Internal distraction	Reading, moving object, handling insect or pet	Binary	0.057	0.054	0	1
Distraction 6. Dining	Includes eating or drinking	Binary	0.022	0.022	0	1
Distraction 7. Other	Smoking, external distraction, personal hygiene, driving-related inattention to forward roadway	Binary	0.106	0.095	0	1
Driving context						
Alignment	Curve (1); tangent (0)	Binary	0.313	0.215	0	1
Lighting	Dawn or dusk (1); day (0)	Binary	0.062	0.058	0	1
Surface condition	Dry (1); wet, icy, or snowy (0)	Binary	0.189	0.154	0	1
Traffic density	Non-free-flow (1); free-flow (0)	Binary	0.247	0.186	0	1
Driving style						
DDDI aggressive driving index	Scale reflects intent to harm	Continuous	11.439	1.414	7	23
DDDI negative emotions index	Scale reflects negative emotions during driving	Continuous	21.193	1.414	11	34
DDDI risky driving index	Scale reflects reckless driving	Continuous	19.281	0.707	12	31
LSI	Scale reflects stress in one's life	Continuous	180.2	45.3	0	560
Driver attribute: driver experience	Number of years with license	Continuous	16.167	7.071	1.5	5

NOTE: SD = standard deviation; min = minimum; max = maximum; GES = general estimates system.

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. Details of the various categories for each environmental factor considered can be seen in Table 3.

Certain driver attributes and personality descriptors were obtained before and after data collection. Years of driving experience were self-reported by all participants on the basis of the number of years they had had a driver's license. The Dula Dangerous Driving Index (DDDI) was used to measure drivers' self-reported likelihood of dangerous driving (*23*). Each DDDI scale (DDDI total, aggressive driving, negative emotional driving, and risky driving) had tests of internal reliability, and evidence of construct validity of the scales as part of initial scale development and testing (*23*). Participants responded to the items on the following Likert scale: A = never, B = rarely, C = sometimes, D = often, and E = always. To quantify DDDI, numerical values were assigned to each response (1 through 5 for A through E, respectively). The higher the score per driver, the more dangerous the driving behavior was considered to be. The Life Stress Index (LSI) questionnaire was used to obtain information about various types of stress and changes that subject drivers may have experienced in the year before the study.

The DDDI and LSI were used to describe aspects of individuals' personalities or current life experiences that may have predisposed them to crash, near crash, or to be involved in a critical incident. These measures were included in the models to provide additional insight about event causality. Other researchers have supported the general validity of seeking to associate these types of descriptors with crash occurrence (*1*). The reliability of personality measures over time has been addressed in at least one study that found reasonable correlation (*24*). Particularly given the recent interest in aggressive driving, it seemed reasonable to explore associations between the DDDI and event odds. Given that the LSI measured a different dimension of stress that could affect driving, it seemed reasonable to include this measure in the models as well.

Table 3 is a list of variable names, definitions, types, data sources, and statistical summaries for run-off-road-related events. All covariates available in the VTTI data set were tested in the analysis. The predictors shown in Table 3 are those that extensive modeling indicated were most consistently associated with event outcomes. Literally hundreds of models were explored to produce the reduced set of predictors in Table 3 (*16*). The predictor variables were divided into four primary groups as shown in Figure 1.

RESULTS

Comparison of Single-Level Models: Effects of Omitting Predictors

Table 4 contains the estimation results obtained by implementing the analysis approach described previously. To reduce the number of models in this study, three models were selected that illustrated the effects of omitting predictors; similar results are illustrated in

TABLE 4 Summary of Estimated Binary Logit Models

	Context-Only Predictors			Context and Driver Attribute Predictors			Context, Driver, and Event Predictors		
Variable	Coeff.	SE	Odds Ratio	Coeff.	SE	Odds Ratio	Coeff.	SE	Odds Ratio
Intercept	–1.16	0.23	NA	0.90	1.17	NA	–1.67	1.81	NA
Event Attributes									
Precipitating Event 1. Loss of control	NA	NA	NA	NA	NA	NA	1.134	1.02	3.11
Precipitating Event 2. Subject over lane line or road edge	NA	NA	NA	NA	NA	NA	2.27	1.00	9.67
Impairment 1. Drowsy, sleepy, asleep, or fatigue	NA	NA	NA	NA	NA	NA	1.57	0.62	4.81
Distraction 1. Wireless device	NA	NA	NA	NA	NA	NA	0.78	0.78	2.18
Distraction 2. Vehicle related	NA	NA	NA	NA	NA	NA	2.22	0.94	9.24
Distraction 3. Passenger related	NA	NA	NA	NA	NA	NA	1.79	0.85	6.01
Distraction 4. Talking, singing, or daydreaming	NA	NA	NA	NA	NA	NA	2.00	1.09	7.36
Distraction 5. Internal distraction	NA	NA	NA	NA	NA	NA	3.09	0.99	21.89
Distraction 6. Dining	NA	NA	NA	NA	NA	NA	1.88	1.31	6.55
Distraction 7. Other	NA	NA	NA	NA	NA	NA	1.25	0.74	3.49
Driving Context									
Alignment 1. Curve	0.64	0.33	1.90	0.48	0.37	1.61	0.93	0.48	2.54
Lighting 1. Dawn or dusk	1.11	0.61	3.02	0.74	0.66	2.10	2.29	0.74	9.91
Surface Condition 1. Wet, icy, or snowy	0.08	0.41	1.08	–0.17	0.46	0.84	0.92	0.62	2.51
Traffic Density 1. Non-free-flow	–1.63	0.55	0.20	–2.00	0.60	0.14	–2.19	0.69	0.11
Driver Attributes									
DDDI aggressive driving index	NA	NA	NA	–0.10	0.04	0.91	–0.12	0.05	0.89
DDDI negative emotions index	NA	NA	NA	–0.10	0.06	0.91	–0.09	0.07	0.92
DDDI risky driving index	NA	NA	NA	0.05	0.08	1.05	0.05	0.09	1.05
Driver experience	NA	NA	NA	–0.03	0.02	0.97	–0.06	0.02	0.94
LSI	NA	NA	NA	0.004	0.001	1.004	0.004	0.002	1.00

NOTE: Coeff. = coefficient; SE = standard error; NA = not applicable. $p = .20$ was used to include variables liberally in the model for exploratory purposes.

other pair-wise comparisons. The binary logit model estimates with context-only predictors are shown in Table 4, including the coefficient mean, standard error, and odds ratio. The discussion of the models focuses on the sign, magnitude, standard deviation, and odds ratio for each predictor. Statistical significance is discussed, but the limitations of the modeling as a result of sample size are recognized.

The first model considered context only. This model is familiar to many road safety analysts, because context variables form the primary variables typically included in a safety performance function, which is fundamental to contemporary road safety management. Alignment and lighting condition indicated increased crash odds on a curve, or at dawn and dusk (odds ratio greater than 1.0). Icy, wet, or snowy conditions increased crash odds, but the effect was small in magnitude (8%) and insignificant at any conventional significance level. As traffic density increased (was non-free-flowing), the likelihood of a run-off-road event decreased.

Changes occurred when driver attributes were added to the model. All parameter values for context predictors decreased, while standard deviation values increased. Collectively, driver attributes themselves had a small effect on crash odds; most odds were less than 10% in magnitude. This was an initial indication that omitted-variable bias was present in this naturalistic event data set.

The situation became even more dramatic once the event attributes were added. Odds ratios were doubled, or tripled, for some context predictors (e.g., the parameter for lighting was 2.29 rather than 1.11; the surface condition changed from 0.08 to 0.92). The odds ratios showed corresponding increases in magnitude. The goodness-of-fit measures changed substantially as well (Table 5): log likelihood and pseudo-R^2 values showed substantial improvement in model fit.

TABLE 5 Goodness-of-Fit Statistics for Estimated Binary Logit Models

Measure	Context-Only Predictors	Context and Driver Attribute Predictors	Context, Driver, and Event Predictors
Log likelihood	−117.80	−103.42	−82.812
Pseudo-R^2	.08	.19	.35

What does this imply for modeling naturalistic events? There is no question that the omission of variables resulted in a substantial bias in parameters; in these data, this was especially true for the context variables. There was less change in driver attributes. Failure to include event characteristics in the model ran the risk of substantial misestimation of the effect of context on event occurrence. The next step was to compare the best single-level logit model (Table 4, last three columns) with a multilevel logit formulation.

Multilevel, Event-Based Model

Table 6 compares the best single-level logit and the multilevel, event-based model. As before, parameter values, standard errors, and odds ratios are included. At Level 1, the event-based data set

TABLE 6 Comparison of Single-Level Logit and Multilevel Random Intercept Models

	Single-Level Logit[a]				Multilevel-Random Intercept[b]			
	Parameter				Parameter			
Variable	Coeff.	SD	Odds Ratio	Sig.	Coeff.	SD	Odds Ratio	Sig.
Intercept	–1.67	1.81	NA		–1.89	2.649	NA	
Precipitating Event 1. Loss of control	1.135	1.02	3.11		1.586	1.267	4.88	
Precipitating Event 2. Subject over lane line or road edge	2.269	1.00	9.67	*	2.948	1.275	19.07	*
Driver Impairment 1. Drowsy, sleepy, asleep, or fatigue	1.571	0.62	4.81	*	2.281	0.82	9.79	*
Distraction 1. Wireless device	0.78	0.78	2.18		1.152	0.91	3.16	
Distraction 2. Vehicle related	2.224	0.94	9.24	*	2.917	1.161	18.49	*
Distraction 3. Passenger related	1.794	0.85	6.01	*	2.454	1.066	11.63	*
Distraction 4. Talking, singing, or daydreaming	1.996	1.09	7.36	**	2.718	1.405	15.15	**
Distraction 5. Internal distraction	3.086	0.99	21.88	*	3.992	1.336	54.16	*
Distraction 6. Dining	1.879	1.31	6.55		2.229	1.675	9.29	
Distraction 7. Other	1.248	0.74	3.48	**	1.63	0.862	5.10	***
Alignment 1. Curve	0.932	0.48	2.54	*	1.005	0.54	2.73	***
Lighting 1. Dawn or dusk	2.293	0.74	9.90	*	2.75	0.855	15.64	*
Surface Condition 1. Wet, icy, or snowy	0.919	0.62	2.51		1.283	0.743	3.61	***
Traffic Density 1. Non-free-flow	–2.19	0.69	0.11	*	–2.916	0.986	0.05	*
DDDI aggressive driving index	–0.12	0.05	0.89	*	–0.168	0.078	0.85	*
DDDI negative emotions index	–0.09	0.07	0.91		–0.114	0.098	0.89	
DDDI risky driving index	0.047	0.09	1.05		0.054	0.083	1.06	
Driver experience	–0.06	0.02	0.94	*	–0.073	0.031	0.93	*
LSI	0.004	0.00	1.00	*	0.005	0.003	1.01	***

[a]Receiver operating characteristic area = 0.8809.
[b]Variance partition coefficient = 27.26%; receiver operating characteristic area = 0.8801.
*Significant (sig.) at $\alpha = .01$, **$\alpha = .05$, and ***$\alpha = .10$.

presents two types of variables: event attributes (occurrences inside the car) and context (occurrences outside the car). Level 2 models driver attributes, which represented the varying effects of driving style (DDDI and LSI) and years of driving experience.

One consistent finding was that the standard errors in the fully specified, single-level model were underestimated compared with those in the multilevel model. This illustrated the concern for multiple observations of individual drivers in the single-level model. It was clear that event-based analyses of naturalistic driving data needed to account for the multiple observations of individual drivers through the use, for example, of Bayesian hierarchical models, which would be one approach.

Precipitating Event 2: The category "subject over lane line or road edge" was statistically significant (the category "loss of control" was not). Although loss of control included various situations in which drivers lost control of the vehicle, most precipitating events in this category were the result of excessive speed or poor road conditions. The subject over lane line or road edge category included movement over the left or right lane line or road edge. Both precipitating event parameters estimates were positive, which indicated that these behaviors were associated with an increased likelihood of a crash or near crash, as compared with a critical incident.

Among the various distractions, those that were internal or vehicle or passenger related were significant, whereas talking, singing, and daydreaming were marginally significant. All had coefficients that were positive in sign, which indicated that they increased the likelihood of a crash or near crash as compared with an incident. Use of a wireless device (at least at the time of data collection) was not significant but generally increased crash or near-crash odds by 3.2 times. Driver impairment that involved drowsiness, sleepiness, and fatigue was also significant and positive (a 9.8 times increase in crash or near-crash odds). The results showed that not all sources of distractions were equally distracting; internal distraction was the most hazardous, and it increased crash probability by roughly 54 times compared with no distraction. These were high odds ratios, but they reflected the increases in odds posed by distractions in the last few seconds before a precipitating event.

Two variables were significant determinants of context: travel in non-free-flowing traffic conditions reduced crash or near-crash odds for run-off-road events, while travel at dawn or dusk substantially increased crash or near-crash odds. Non-free-flowing traffic density reduced the likelihood of a run-off-road crash, or near crash, by 95%, which increased the likelihood of a comparable critical incident. Dawn and dusk lighting conditions increased the run-off-road crash or near-crash likelihood by 15.6 times. Curve presence and surface conditions that were wet, icy, or snowy were both marginally significant and positive. They increased the crash or near-crash odds by 2.7 times and 3.6 times, respectively. These results showed a strong dependency between context and distraction. Their interaction should be accounted for in any future use of naturalistic data to study distraction.

Several driver-level variables were significant. Driving experience was significant in the differentiation of event severity: those with more experience driving (measured as self-reported years of driving) had a reduced likelihood of a crash or near crash compared with an incident. The addition of 1 year of driving experience decreased the crash or near-crash odds by 7%. The effect of life stress was marginally significant and positive. The model indicated that aggressive drivers or ones that had negative emotions as they drove (on the basis of DDDI, aggressive driving, and negative emotions scores) were more likely to become involved in critical incidents (i.e., they were less likely to have crashes or near crashes). This finding was a difficult one to interpret; additional studies are needed to either support or refute this result. In particular, the validity of the DDDI is in need of additional exploration to permit more confidence in the interpretation of these findings.

CONCLUSIONS AND RECOMMENDATIONS

This paper reports on results from studies of the modeling of naturalistic, event-based driving data, with a specific focus on omitted-variable bias and the application of multilevel models to the data. On the basis of data that concerned road departure events from the VTTI 100-car study, the team used a series of logit models to estimate the probability of a crash or near crash as opposed to a critical incident. Among the findings were the following:

1. Omitted-variable bias was apparent in the models, particularly for context variables such as road surface condition and lighting. Odds ratios tripled for lighting and more than doubled for road surface condition.
2. Driver-related variables (driving experience and a set of measures of driving risk propensity) changed a small amount; omitted-variable bias was much less apparent with these predictors.
3. There was little question that failure to include predictors from each of three variable types (driver, event, and context) had serious implications for variable bias, and especially for context.
4. Multilevel models revealed that heterogeneity was a problem in estimating event-based models because some drivers had been involved in multiple events, which needed to be recognized in the model formulation. The team chose multilevel models to address this issue, but other methods may also be used.
5. The particular advantage of the multilevel approach is that it uses a structure that reflects how driver decisions are made: drivers with particular characteristics (one level) find themselves in certain contexts in which they execute specific driving maneuvers (second level), which lead to certain outcomes.

It is critically important that omitted-variable bias be identified with naturalistic data. The primary advantage of naturalistic data is that factors not previously observed or estimated through use of judgment are now observable with, it is hoped, a high degree of accuracy and reliability. It would be a shame to give up that accuracy to a poor model specification. Tests with additional data sets are needed to provide verification, but the need to include context variables in event-based analysis seems strong.

Some may be concerned that the models in this study did not contain nonevent (baseline) events. Inclusion of baseline events would allow an estimation of the odds of a crash as compared with baseline noncrashes (a rough estimate of exposure). This area is ripe for further research, but nonevents are costly to obtain, particularly if an array of typical predictors is to be included in a data set. These data are obtained at random from a large driving file. It remains an open question as to how many observations will be needed to address random effects that may occur.

The models developed in this study compared three types of events. The comparison of crashes and incidents could be thought of as a search to identify factors that differentiated crash events from these other events: a surrogate event for a crash. Although it is beyond the scope of this paper, a recent report discussed the application of event-based analyses to surrogate testing (*16*). Categorical models can be used to identify events characterized as critical incidents, which can be used as surrogates for crashes; the model can help identify the closeness of a critical incident event to a crash event. Similar statistical comparisons can be made with crashes and near crashes. What are required are sufficient data and a good model specification.

REFERENCES

1. Shinar, D. *Traffic Safety and Human Behavior.* Elsevier, Amsterdam, Netherlands, 2007.
2. Jovanis, P. P., and H.-L. Chang. Modeling the Relationship of Accidents to Miles Traveled. In *Transportation Research Record 1068,* TRB, National Research Council, Washington, D.C., 1986, pp. 42–51.
3. Shankar, V. N., F. L. Mannering, and W. Barfield. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, No. 3, 1995, pp. 371–398.
4. Lord, D., and F. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A,* Vol. 44, 2010, pp. 291–305.
5. *Highway Safety Manual.* AASHTO, Washington, D.C., 2010.
6. 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. *The 100-Car Naturalistic Driving Study, Phase II: Results of the 100-Car Field Experiment.* HS 810 593. Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, Va.; NHTSA, U.S. Department of Transportation, Washington, D.C., 2005.
7. Dingus, T. A., S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, and J. Sudweeks. *The 100-Car Naturalistic Driving Study. Phase II: Results of the 100-Car Field Experiment,* HS 810 593. Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, Va.; NHTSA, U.S. Department of Transportation, Washington, D.C., 2006.
8. Hanowski, R. J., M. A. Perez, and T. A. Dingus. Driver Distraction in Long-Haul Truck Drivers. *Transportation Research Part F,* Vol. 8, No. 6, 2005, pp. 441–458.
9. Hanowski, R. J., J. Hickman, M. C. Fumero, R. L. Olson, and T. A. Dingus. The Sleep of Commercial Vehicle Drivers Under the 2003 Hours-of-Service Regulations. *Accident Analysis and Prevention,* Vol. 39, No. 6, 2007, pp. 1140–1145.
10. Hanowski, R. J., J. Hickman, W. W. Wierwille, and A. Keisler. A Descriptive Analysis of Light Vehicle-Heavy Vehicle Interactions Using in Situ Driving Data. *Accident Analysis and Prevention,* Vol. 39, No. 1, 2007, pp. 169–179.
11. Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, Va. http://www.vtti.vt.edu/casr-research-details.html#teen-nds. Accessed July 22, 2010.
12. Bogard, S., P. Fancher, R. Ervin, M. Hagan, and Z. Bareket. *Intelligent Cruise Control Field Operational Test.* Final report, Vol. III. NHTSA, U.S. Department of Transportation, Washington, D.C., 1998.
13. LeBlanc, D., J. Sayer, C. Winkler, S. Bogard, J. Devonshire, M. Mefford, M. Hagan, Z. Bareket, R. Goodsell, and T. Gordon. *Road Departure Crash Warning System Field Operational Test.* Final report. U.S. Department of Transportation, Washington, D.C., 2006.
14. University of Michigan Transportation Research Institute and General Motors Research and Development Center. *Automotive Collision Avoidance System Field Operational Test Report: Methodology and Results.* Final research report. DTNH22-99-H-07019. NHTSA, U.S. Department of Transportation, Washington, D.C., 2005.
15. Shankar, V. N., P. P. Jovanis, J. Aguero-Valverde, and F. Gross. Analysis of Naturalistic Driving Data: Prospective View on Methodological Paradigms. In *Transportation Research Record: Journal of the Transportation Research Board, No.* 2061, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 1–8.
16. Jovanis, P. P., V. Shankar, J. Aguero-Valverde, K. Wu, and A. Greenstein. *SHRP 2 Report S-01b: Analysis of Existing Data: Prospective Views on Methodological Paradigms.* Draft final report. Larson Transportation Institute, Pennsylvania State University, University Park; Transportation Research Board of the National Academies, Washington, D.C., 2010.
17. Washington, S., M. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman & Hall/CRC, Boca Raton, Fla., 2003.
18. Rasbash, J., S. Fiona, W. Browne, and B. Prosser. *A User's Guide to MLwiN.* Center for Multilevel Modeling, University of Bristol, Avon, United Kingdom, 2008.
19. Snijders, T. A. B., and R. J. Bosker. *Multilevel Analysis.* Sage Publications, Newbury Park, Calif., 1988.
20. Swets, J. A. Measuring the Accuracy of Diagnostic Systems. *Science,* Vol. 240, pp. 1285–1293, 1988.
21. Agresti, A. *Categorical Data Analysis,* 2nd ed. Wiley-Interscience, Inc., Hoboken, N.J., 2002.
22. Neale, V. L., T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman. *Overview of the 100-Car Naturalistic Study and Findings.* VTRC-05-0400. Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, Va.; NHTSA, U.S. Department of Transportation, Washington, D.C., 2005.
23. Dula, C. S., and M. E. Ballard. Development and Evaluation of a Measure of Dangerous, Aggressive, Negative Emotional and Risky Driving. *Journal of Applied Social Psychology,* Vol. 33, No. 2, 2003, pp. 263–282.
24. Ozkan, T. O., T. Lajunen, and H. Summala. Driver Behaviour Questionnaire: A Follow-Up Study. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 386–395.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.