ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Statistical Methods for Naturalistic Driving Studies

Feng Guo

Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061–0439, USA; email: feng.guo@vt.edu

Annu. Rev. Stat. Appl. 2019. 6:309–28

First published as a Review in Advance on November 12, 2018

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-030718-105153

Keywords

naturalistic driving study, traffic safety, driver behavior, distraction, case-cohort, case-crossover, generalized linear models, recurrent event models

Abstract

The naturalistic driving study (NDS) is an innovative research method characterized by the continuous recording of driving information using advanced instrumentation under real-world driving conditions. NDSs provide opportunities to assess driving risks that are difficult to evaluate using traditional crash database or experimental methods. NDS findings have profound impacts on driving safety research, safety countermeasures development, and public policy. NDSs also come with attendant challenges to statistical analysis, however, due to the sheer volume of data collected, complex structure, and high cost associated with information extraction. This article reviews statistical and analytical methods for working with NDS data. Topics include the characteristics of NDSs; NDS data components; and epidemiological approaches for video-based risk modeling, including case-cohort and case-crossover study designs, logistic models, Poisson models, and recurrent event models. The article also discusses several key issues related to NDS analysis, such as crash surrogates and alternative reference exposure levels.

INTRODUCTION

Driving is an essential part of modern life, allowing for a high level of convenient mobility for both passengers and commodity shipment. However, high dependency on automobile travel is also associated with adverse effects, such as pollution, congestion, and crashes. In the United States, traffic crashes cause 30,000 to 40,000 fatalities annually and are the number one cause of death among 8- to 24-year-olds (NHTSA 2015). Identifying the cause of crashes and developing safety countermeasures has been a major focus of government agencies, academia, and industry for decades. Many factors contribute to crashes, including the attributes of vehicles, traffic conditions, road infrastructure, driving environment, and drivers. Crash databases have traditionally been the major resource for evaluating crash risk, including, for example, the Crashworthiness Data System, the General Estimation System (GES), the Fatality Analysis Reporting System (FARS), and various crash databases maintained by individual states. However, while these databases include a large number of crashes—FARS contains a complete census of all fatal crashes—they are based on postcrash investigation where information bias could be severe.

Driver behavior is a major contributing factor to crashes. Previous studies have shown that driver behavior could contribute to more than 90% of crashes (Dingus et al. 2016, Treat et al. 1979). A recent study by the National Highway Traffic Safety Administration (NHTSA) indicated that human error was the critical reason for 93% of crashes (NHTSA 2013). One challenge in evaluating crash risk is the difficulty in collecting measurements of rapidly changing risk factors, including driver behavior, traffic condition, and vehicle condition at any given moment. The challenge comes from both the demand for high-precision information and the difficulty of generating specific driving scenarios. Driving simulators and test track experiments can control the driving environment and record detailed driver behavior information, but they suffer from several limitations. A driver's behavior in these conditions is likely to differ considerably from what it would be in real-life driving scenarios and therefore might not provide realistic information. Another major obstacle is that crashes cannot be readily observed in these experimental settings, making crash risk estimation difficult if not impossible.

Large scale naturalistic driving studies (NDSs) fill in the gaps between crash databases and driving experiments (Dingus et al. 2006). An NDS is characterized by its continuous recording of driving data under real-life (i.e., naturalistic) driving conditions for an extended period. The key technology required is unobtrusive instrumentation installed on a driver's vehicle to continuously collect data. Typical instrumentation includes cameras to collect multiple views, a GPS, speedometers, multidimensional accelerometers, an ambient lighting meter, and access to the vehicle's network information. There are typically no specific driving instructions or requirements given to NDS participants. Rather, participants drive the instrumented vehicles as they normally would, and the recorded data thus reflect natural driving behavior and environments.

The past decade has witnessed a substantial increase in NDSs worldwide. Large scale general data collection studies to date include the second Strategic Highway Research Program (SHRP 2) NDS (with more than 3,400 participants), the 100-Car NDS, and Europe's UDRIVE NDS (Dingus et al. 2006, 2015; Eenink et al. 2014). NDSs with specific target populations have also been conducted, including the 40-Teen Naturalistic Driving Study, the Older Driver Fitness-to-Drive NDS, and the Commercial Truck Driver Naturalistic Driving Study (Guo et al. 2015, Klauer et al. 2014, Olson et al. 2009). NDSs provide valuable insight into crash causation and factors that substantially increase crash risk, such as eye-off driving tasks or texting while driving. NDS findings have had direct impacts on legislation regulating texting while driving, guidelines related to in-vehicle behavior, and the development of safety countermeasures (NHTSA 2012).

Along with the many benefits an NDS has to offer comes the challenge of working with the type and quantity of data collected. The sheer volume and the complex data structure present challenges for both data processing and statistical analysis. Efficiently extracting information from NDS data is typically beyond the capacity of individual researchers, making database programmers and video reduction teams essential for the success of a study. This requires the involvement of statisticians in the early stages of planning to provide guidance on study design, data extraction, sample size, and analysis. Despite the challenges that the complex NDS data structure presents, it also provides opportunities to develop novel statistical methods. This article provides an overview of NDSs and the commonly used statistical methods, as well as some of the associated challenges of modeling NDS data.

NATURALISTIC DRIVING STUDIES

The development of NDS methods is driven by the advancement of in-vehicle data collection technology. There are several key requirements for the data acquisition system (DAS) used in an NDS. The DAS should be small and unobtrusive to minimize any interference with the driver's operation of the vehicle, there should be sufficient data storage space to allow for extended periods of data collection, and the installation labor and time should be within certain a limit to make large-scale vehicle instrumentation feasible. Multiple types of DAS have been used in different NDSs worldwide. Notably, the Virginia Tech Transportation Institute is a pioneer in DAS development, having developed DASs for the first large-scale NDS, the 100-Car NDS, and the SHRP 2 NDS, which is the largest NDS conducted to date (Dingus et al. 2006, 2015).

Videos contain some of the most important information from NDSs. The next generation (NextGen) DAS used in the SHRP 2 NDS included four camera views: a front view, a driver's face view, an over-the-shoulder view, and a rear view, as shown in **Figure 1**. These views were strategically selected to cover critical driving environment elements and driver behavior. Video setup required balancing resolution needs with data storage requirements, which was achieved by recording only the front view in color and by varying the size of camera views. The NextGen DAS was able to store up to 6 months of data under a normal driving load before it was necessary to change the onboard hard drive.

In addition to videos, the DAS also recorded dozens of continuous driving variables, including information from a GPS, radars, speedometers, multidimensional accelerometers, an ambient lighting meter, and an alcohol sensor. With the cooperation of vehicle manufacturers, the NextGen DAS could also access vehicle network information, such as steering wheel angle, accelerator pedal position, turn signal, etc. The frequency of data was asynchronous—video, acceleration, and radar were typically collected at a higher frequency (e.g., 10 Hz or 15 Hz), while GPS and speed derived from GPS were collected at 1 Hz. Specialized database processing was required to match the variables by time stamp.

While continuous driving data with video recordings are generally considered key NDS characteristics, NDSs can also be defined as any study without specific instructions as to how participants should drive. Instead of continuous data collection, some data collection systems, such as the DriveCam system, only record data when a kinematic event is triggered, e.g., deceleration rate exceeds a certain threshold (Hickman et al. 2010). Other NDSs, such as the euroFOT study, may not include video recordings. Regardless of how or what data were collected, the common characteristic of all these NDSs is that all driving was done during regular day-to-day activities instead of for research purposes, and, as a result, the data reflect real-life, naturalistic driving.

NDSs typically follow the prospective cohort study design, in which participants are recruited at the beginning of the study and exposure and cases information are collected during the study.

Figure 1

Second Strategic Highway Research Program (SHRP 2) naturalistic driving study quad image of video views. Adapted with permission from Dingus et al. (2015).

The NDS thus inherits the characteristics of a cohort study, which includes high associated costs that limit the sample size, the ability to answer multiple research questions, and the ability to collect a large amount of data. The last of these characteristics also makes data management and analysis challenging. For example, the SHRP 2 NDS includes 5.4 million trips, more than 1 million hours of continuous driving data, approximately 35 million driving miles, and more than 2 petabytes of data. This amount of data requires novel epidemiological and statistical methods to model the NDS data according to specific research questions of interest.

Naturalistic Driving Studies, Crash Databases, and Empirical Data Collection

An NDS database provides a unique perspective on driving compared with traditional databases and fills the gap between a traditional epidemiological data collection approach and empirical data collection methods, as illustrated in **Figure 2** (Dingus et al. 2006). Epidemiological data collection refers to the traditional police report–based crash database or postcrash investigation–generated crash database. Epidemiological collections include FARS, which provides a census of fatal traffic crashes in the United States; the GES, which provides a weighted estimation of annual crashes; and most state-maintained crash databases. The crash databases are limited by the amount of information available and the quality of the data collection. For instance, police reports often do not include information crucial for crash causation evaluation. One key drawback to these types of traditional crash databases is the lack of precise information about the vehicle, environment, and driver during the moments prior to a crash, which makes it difficult to evaluate the risk level of the driver's behavior and other time-variant risk factors with transient effects.

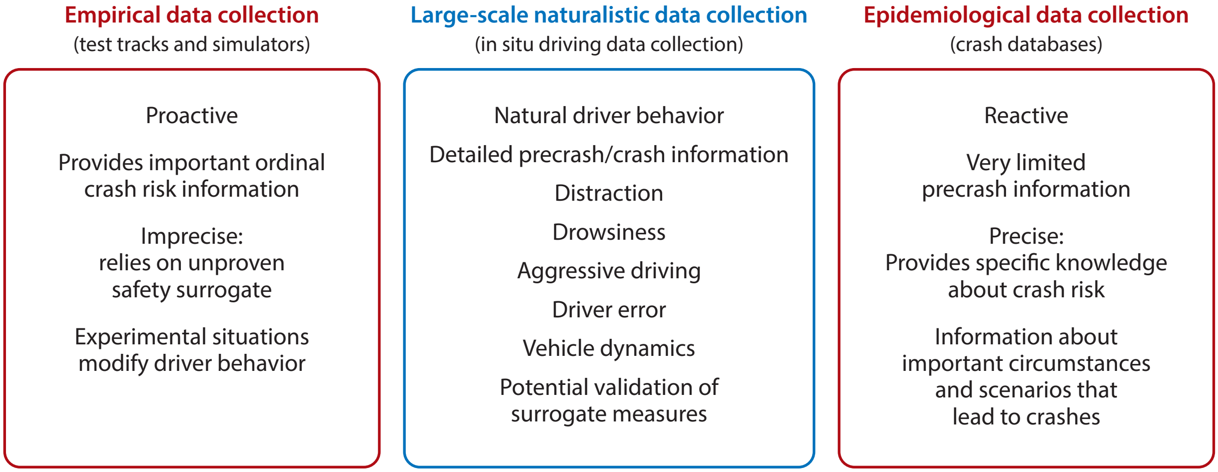

Figure 2

Comparison of empirical, naturalistic, and epidemiological methods in driving safety research. Adapted with permission from Dingus et al. (2006).

Empirical data collection, the collection of driving data in a controlled environment such as a test track or driving simulator, includes the recording of precise information by sophisticated instruments on the test vehicle or simulator. The main issue with empirical data collection is that the driver's behavior in these controlled environments might not be the same as it would be in real-life driving scenarios, and the simulated scenario is therefore not likely to reflect true driving behavior. In addition, crashes cannot be readily observed in an empirical study, and the sample sizes for this type of study are typically small. In comparison with epidemiological or empirical studies, large scale NDSs can capture precise information both during safety-critical events (e.g., crashes, near-crashes, crash-relevant conflicts, and unintentional lane deviations) and under normal driving conditions. The objectively collected data eliminate the potential information bias that may be present in typical crash databases and truly reflect real-life driver behavior and the driving environment.

Data Fusion

Large scale NDSs are major data collection efforts and are commonly accompanied by a number of other ancillary data collections. For example, the SHRP 2 NDS included multiple data collections on drivers, vehicles, cell phone use, and roadways, as illustrated in **Figure 3** (Hankey et al. 2016). Each participant took a questionnaire survey on demographics, driving history, medical conditions, crash history, and personality characteristics (Dingus et al. 2015). The study also collected detailed vehicle information, such as make, model, and condition, as well as safety and entertainment systems (Hankey et al. 2016). To enable the evaluation of road infrastructure and driver behavior under various road conditions, a study was conducted to collect detailed road characteristics, including alignment, lighting, guardrails, grade, and number of lanes (Smadi et al. 2015). A subset of drivers also provided detailed cell phone records from carriers during the study period (Cook et al. 2015). The collected raw data were processed to extract essential information. Contained within the database, for example, are the trip summary data, which include trip information such as starting time, trip duration, and average speed. One important derived data set is the event data set,

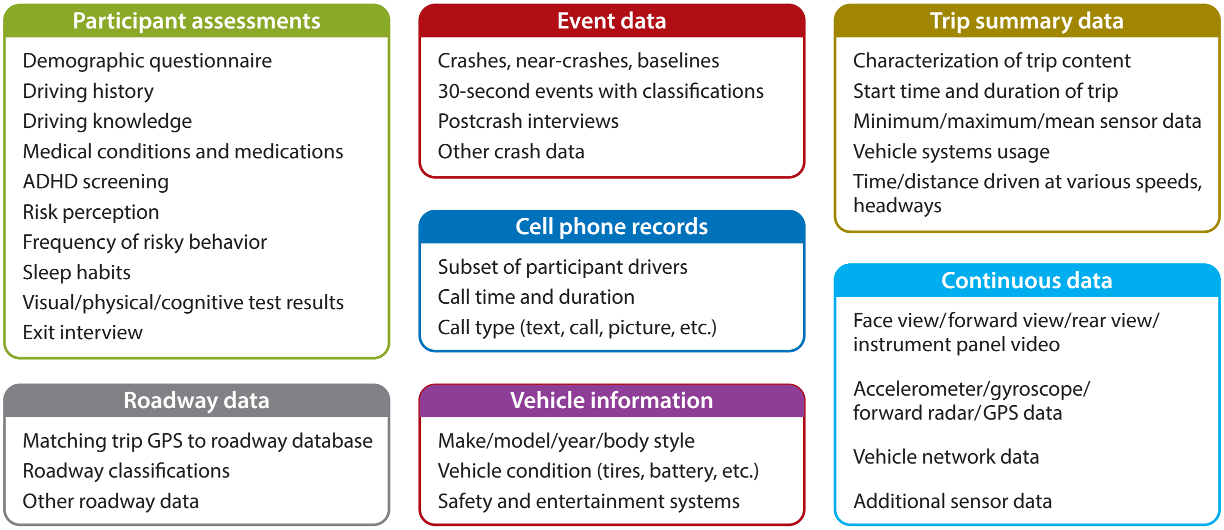

Figure 3

Representation of the data categories collected in the second Strategic Highway Research Program (SHRP 2) naturalistic driving study. Adapted with permission from Hankey et al. (2016).

which includes information such as driver distraction, traffic conditions, environmental conditions, and postcrash interview data from crashes, near-crashes, and control periods of driving. Many variables were extracted from raw data by visually examining the videos for crashes and control periods of driving (Hankey et al. 2016). The control periods were sampled according to the case-cohort study design and are discussed in the following section. Linking and merging various data sources results in richer information than relying on a single data source and is essential for certain types of analysis, such as the evaluation of cell phone use near intersections. Fitch et al. (2013), for instance, merged the SHRP 2 NDS driving data and cell phone records to get an estimation of cell phone use prevalence and associated driving risk.

Crashes and Crash Surrogates

Crash analysis is considered the gold standard in evaluating traffic safety (AASHTO 2010). However, relative to the number of overall vehicle miles traveled, crashes are rare events; the fatality rate in the United States is 1.13 per 100 million vehicle miles traveled, and the property-damage-only crash rate for passenger vehicles is 313 per 100 million miles traveled (NHTSA 2017). The relatively rare occurrence of crashes makes it infeasible to assess safety in certain cases; for example, it is estimated that hundreds of years of data might be needed to confirm that autonomous vehicles could lead to a lower fatality rate than human drivers (Kalra & Paddock 2016). Identifying all crashes in an NDS is ideal for achieving maximum statistical power, and to this end, sophisticated, multistep algorithms have been developed for crash identification (Hankey et al. 2016). The kinematic variables, including longitudinal and lateral acceleration, yaw rate, swerve, and active safety system activations, are examined first, followed by a visual confirmation of videos for the driving segments with high kinematic values. The outputs of this process typically contain the majority of the crashes. However, the relatively small number of participants in any NDS necessarily limits the number of crashes that can be observed, and thus crash surrogates are commonly used in NDS risk evaluation (Guo et al. 2010a, Klauer et al. 2014).

Caution should be used when doing so, however, as discrepancies between crash surrogates and crashes could potentially lead to unexpected results (Knipling 2015). One key point is that the validity of a crash surrogate highly depends on the research objective. Surrogates differ from crashes by definition, and a general statement regarding whether a surrogate is valid without specific context is impossible. The proper way to evaluate a surrogate's validity is, for a given research objective, to determine whether the candidate crash surrogate can provide useful information to answer the research question. A surrogate may be valid for one specific purpose but not for others.

Surrogates have long been used in traffic safety evaluation. Studies using traffic conflicts as a measure of crash potential date back to the 1960s (Perkins & Harris 1968). The Federal Highway Administration developed the "Traffic Conflict Techniques for Safety and Operations" guidelines to diagnose safety and operational problems and evaluate the effectiveness of safety countermeasures (Parker & Zegeer 1989a,b). Advances in video analysis technology allow objective evaluation of surrogate measures, typically based on spatial and temporal proximity. Proximity-based surrogates, most of which utilize onboard sensors (e.g., radar and accelerometer devices) or onsite video monitors to record the kinematic and spatial information required for quantitative data (Topp 1996), include time to collision (Behbahani 2015, Minderhoud & Bovy 2001), postencroachment time (Cooper 1984, Gettman & Head 2003), gap time (Gettman & Head 2003), headway (Vogel 2003), time advantage (Laureshyn et al. 2010), braking time (Lu et al. 2012), proportion of stopping distance (Gettman & Head 2003), and lateral distance to departure (Tarko 2012). Surrogates benefit safety research by providing a low-cost, quickly executable alternative for the evaluation of safety, bypassing the long waiting period that would be needed to observe a large number of crashes in the real world. For simulator and test track studies, crashes are rarely observed and surrogates enable the assessment of driving risk.

The high-resolution data collected via an NDS provide a rich source for crash surrogate measures. Various crash surrogates have been used, such as near-crashes, crash-relevant conflicts, unintentional lane deviations, and hard braking/accelerating (Guo et al. 2010a,b; Klauer et al. 2014; Olson et al. 2009; Simons-Morton et al. 2013).

A substantial amount of research has been conducted to evaluate the validity of surrogates. Heinrich's Triangle is one of the key concepts in surrogate research (Heinrich et al. 1980). The premise of Heinrich's Triangle is that less severe events occur more frequently than severe events; more importantly, the frequency of severe events can be reduced by reducing the frequency of less severe events. Heinrich's Triangle phenomenon is commonly observed in NDS data. For example, the ratio of crashes to near-crashes and the ratio of near-crashes to critical incidents in the 100-Car NDS were both roughly 1:10 (**Figure 4**). Inclusion of near-crashes and critical

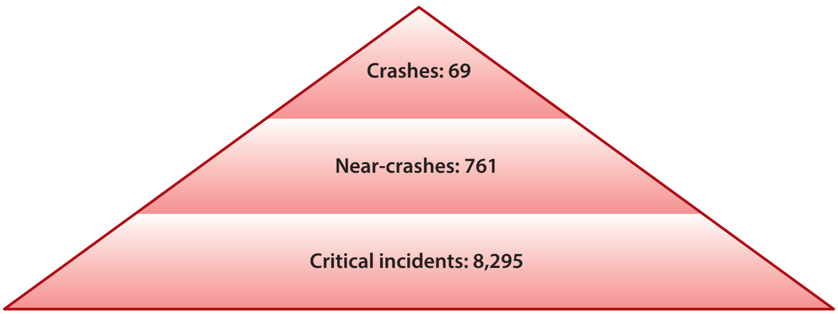

Figure 4

Heinrich's Triangle, 100-Car Naturalistic Driving Study crash, near-crash, and critical incident frequency.

incidents can greatly increase sample size and make driving risk evaluation possible. The critical issue is whether surrogates provide useful information on crashes for the specific purpose of the study.

Guo et al. (2010a) evaluated the validity of using near-crashes to estimate the risk associated with driver distraction. Two principles were proposed: (*a*) Near-crashes have similar causal risk factors as crashes, and (*b*) the ratio between near-crashes and crashes is constant under different conditions. The study showed that when these conditions were satisfied, the odds ratio (OR) under a case-cohort study design was identical when using crashes alone or when using both crashes and near-crashes. An empirical study of the 100-Car study data showed that there was no substantial difference in driver distraction between crashes and near-crashes, though driver reaction differed considerably. There is a strong positive association between the frequency of crashes and near-crashes under various environmental conditions. Finally, using near-crashes seems to generate lower ORs than using crashes alone. The findings of this study were consistent with the results from the large-scale SHRP 2 NDS using crashes alone (Dingus et al. 2016).

Wu & Jovanis (2013) proposed a multistage framework to define, screen, and identify surrogate events and concluded that surrogates can help identify potential countermeasures to improve safety. Jonasson & Rootzén (2014) used extreme value theory to examine the relationship between rear-end crashes and near-crashes. Their findings indicated potential bias but, interestingly, also suggested that near-crashes may be more useful than crashes for improving traffic safety.

As previously noted, the validity of a surrogate measure depends on the study's objectives. For example, certain types of safety-critical events and high G-force events typically provide limited information on the risk of driver behavior; however, studies have shown that they add valuable information in predicting high-risk drivers (Guo & Fang 2013, Simons-Morton et al. 2012). Kinematic events data have been used by industry and researchers to identify high-risk drivers or improve driver behavior. For example, unintentional lane deviation has been shown to be closely related to fatigue-related driving risk (Hanowski et al. 2008, van Dongen et al. 2010).

There are also concerns about using crash surrogates. Knipling (2015) argued that crash surrogates could yield misleading results. The issue is that severe crashes, such as fatal crashes and injury crashes, cause the majority of damages. If surrogates and severe crashes do not share similar causal factors, findings based on surrogates would lead to biased results. The heterogeneity of crashes adds additional challenges to the selection of surrogates.

There are a number of other challenges in crash surrogate research. There is inconsistency in the definition of crash surrogates; for instance, the operational definition of near-crashes varies for different NDS projects. Furthermore, the association between crash and surrogate frequency varies, with some studies reporting strong correlations between crash frequency and crash surrogate frequency and other studies reporting weak correlations (Glauz et al. 1985, Guo et al. 2010b, Wu & Jovanis 2013).

In summary, crash surrogates enable safety research to advance when there is not a sufficient number of crashes available. The validity of crash surrogates depends heavily on the purpose of the research, and caution should be used when selecting appropriate surrogate measures as well as when interpreting the results. When properly used, surrogates can be beneficial to traffic safety research and provide insight into safety issues when the number of crashes is limited.

STATISTICAL MODELS FOR EVALUATING TIME-VARIANT RISK FACTORS

Time-variant risk factors with transient effects, such as driver distraction, eye-off driving tasks, fatigue, driving errors, or certain environmental factors, could contribute to more than 90% of total crashes (NHTSA 2013). The status of these factors changes constantly and rapidly, often in terms of seconds, with crash risk only elevating when they are present. Accordingly, it is important to accurately identify when and where these risk factors occur. The major challenge in accomplishing this is that many factors typically must be identified from videos. Automatic video recognition algorithms have been explored, but the accuracy of results often does not meet analysis requirements (Smith et al. 2017). Due to the complexity of driver behavior and the driving environment, a visual examination of recorded videos by trained data reductionists is still the primary method for extracting time-variant risk factors (Hankey et al. 2016).

Due to the sheer number of recorded video hours, it is not feasible to manually review all the videos in an NDS data set. Data reduction thus typically begins by first identifying crashes or surrogates, as discussed previously. A detailed data reduction protocol is then employed to extract driver behavior, environmental conditions, and traffic information moments prior to the event, typically within a 6-second window (Dingus et al. 2016, Klauer et al. 2014).

For estimating the risk associated with time-variant risk factors, it is necessary to compare exposure to risk factors preceding crashes to exposure to those same factors during normal driving conditions. This is analogous to the case and control in case-control study design. A control in this context is a short driving segment that represents the exposure status of risk factors under normal, non-safety-critical driving conditions. It should be noted that in the context of NDS research, the sampled normal driving segments have been traditionally called baselines, control periods of driving, control driving segments, and baseline exposure segments (Guo et al. 2017; Klauer et al. 2006, 2014; Victor et al. 2015).

Control periods of driving should be sufficiently short that the exposure status can be considered as constant within the window. To make the data compatible, the duration of control periods should be comparable to the duration of data reduction for crashes or crash surrogates (typically 6 seconds). The critical issue is how the control periods should be selected. Two popular epidemiology approaches, case-cohort and case-crossover study designs, are commonly used in selecting control periods and evaluating time-variant risk factors using NDS data (Dingus et al. 2016, Guo et al. 2017, Klauer et al. 2014).

Case-Cohort Random Sampling Design

The case-cohort design is a hybrid design combining the characteristics of both cohort and case-control study designs, in which cases and controls are drawn from within a prospective study (Prentice 1986). It is efficient in studies where raw information has been recorded for a large cohort but eliciting exposure information is difficult, which is exactly the case for evaluating time-variant risk factors in an NDS data set. The cases—the crashes or surrogates—are identified through a separate procedure (Hankey et al. 2016), and the case-cohort design focuses on how controls should be selected.

A commonly used case-cohort design uses random sampling stratified by individual participants' driving exposure (Guo & Hankey 2009, Hankey et al. 2016). Under this study design, control periods are randomly selected for the continuous driving data. The controls are stratified by drivers, and the number of controls for each driver is proportional to the valid moving hours or miles traveled by the driver. The stratification ensures that all participants contribute to the control pool.

The random sampling scheme is illustrated in **Figure 5**. The total driving time/miles under exposed and unexposed conditions is represented by horizontal bars. Frequencies of crashes that occurred under exposed and unexposed conditions are denoted by A and C, respectively. The frequencies of randomly selected control periods are denoted by B and D. The total number of

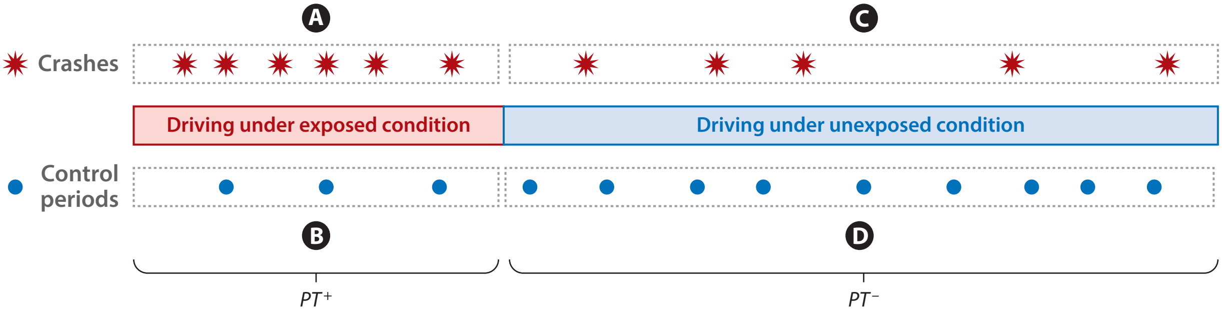

Figure 5

Case-cohort design. A, number of crashes under exposed condition; B, number of control periods under exposed condition; C, number of crashes under unexposed condition; D, number of control periods under unexposed condition; PT^{+}, duration of driving under exposed condition; PT^{-}, duration of driving under unexposed condition.

cases (crashes or crash surrogates), $(A + C)$, and control periods, $(B + D)$, are predetermined. The corresponding contingency table is shown in **Table 1**.

Assume the total exposed duration for a risk factor is PT^{+} and the unexposed duration is PT^{-}. Under this random sampling scheme, the probability that a control period of driving falls in the exposed status (PT^{+}) is proportional to the relative duration; i.e.,

$$p_{\text{exposed}} = \text{Prob(A control period is under exposed condition)} = \frac{PT^{+}}{(PT^{+}) + (PT^{-})}.$$

The expected number of exposed control periods is

$$\mathrm{E}[B] = N_{\text{control}}{}^{*} p_{\text{exposed}} = N_{\text{control}}{}^{*} \frac{PT^{+}}{(PT^{+}) + (PT^{-})},$$

where $N_{\text{control}} = B + D$ is the prespecified total number of control periods. Similarly, the expected number of unexposed control periods is

$$\mathrm{E}[D] = N_{\text{control}}{}^{*} \frac{PT^{-}}{(PT^{+}) + (PT^{-})}.$$

The ratio between B and D is an approximation for the ratio of exposure duration:

$$\frac{B}{D} \approx \frac{N_{\text{control}}{}^{*} \frac{PT^{+}}{(PT^{+})+(PT^{-})}}{N_{\text{control}}{}^{*} \frac{PT^{-}}{(PT^{+})+(PT^{-})}} = \frac{PT^{+}}{PT^{-}}.$$

Table 1 Contingency table for Figure 5

	E^{+}	E^{-}	Total
Cases	A	C	$A + C$
Controls	B	D	$B + D$
Total	$A + B$	$C + D$	$A + B + C + D$

A, number of crashes under exposed condition; B, number of control periods under exposed condition; C, number of crashes under unexposed condition; D, number of control periods under unexposed condition; E^{+}, exposed condition; E^{-}, unexposed condition.

Therefore, the OR is an approximation for the rate ratio as shown below.

$$\mathrm{OR} = \frac{AD}{BC} = \frac{\frac{A}{C}}{\frac{B}{D}} \approx \frac{\frac{A}{C}}{\frac{PT^+}{PT^-}} = \frac{\frac{A}{PT^+}}{\frac{C}{PT^-}} = \frac{\text{crash rate under exposed condition}}{\text{crash rate under unexposed condition}}.$$

Note that this approximation does not rely on the rare event assumption, as is commonly required for OR approximation for risk ratio. One major advantage of the random sampling scheme is that the control periods represent general driving conditions and can be used to evaluate exposure and ORs for any time-variant risk factors.

The random control driving periods also provide an opportunity to evaluate the prevalence of a risk factor as the percentage of exposed control periods; i.e., the prevalence of a risk factor is $B/(B+D)$. This approach has been used to estimate the prevalence of certain risk factors that are difficult to collect (Dingus et al. 2016). This study design is one of the most commonly used NDS designs and has been used in many high impact studies (Dingus et al. 2016, Guo et al. 2017, Klauer et al. 2014, Olson et al. 2009). As a major subset of the SHRP 2 NDS data set, approximately 20,000 control periods were reduced according to this design and made available for public use (**https://insight.shrp2nds.us/**) (Hankey et al. 2016).

The analysis of the case-cohort data can follow standard contingency table analysis or logistic regression models. Logistic regression models are preferred, in general, due to their flexibility in incorporating correlation among observations and controlling potential confounding factors via regression models. The cases and controls in a case-cohort study are inherently correlated. One driver will be likely to have multiple controls and cases. In addition, data are commonly collected from multiple sites. For example, the SHRP 2 NDS included six data collection sites, and the drastic difference in weather, traffic, and driver population among sites could lead to correlation among drivers from the same area. Mixed effect logistics regression can accommodate such data structures.

The setup for a logistic regression for evaluating the relative risk of a time-variant factor is described as follows. Define a binary random variable Y_{ij} representing the outcome of the jth event of driver i such that

$$Y_{ij} = \begin{cases} 1 & \text{Crash} \\ 0 & \text{Control period of driving} \end{cases}, \; i = 1, \ldots, I; \; j = 1, \; \ldots, \; J_i,$$

where I is the number of drivers and J_i is the number of observations for driver i. Assume Y_{ij} follows a Bernoulli distribution with the probability of being a crash as p_{ij}; i.e.,

$$Y_{ij} = \text{Bernoulli}\left(p_{ij}\right).$$

The parameter p_{ij} is specified through the conditional expectation of Y_{ij} given a random effect term; i.e., $p_{ij} = \mathrm{E}[Y_{ij}|\boldsymbol{u}_i]$, where $\boldsymbol{u}_i$ is the random effect. The p_{ij} is connected with a set of covariates with a logit link function,

$$\text{logit}\left(p_{ij}\right) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \boldsymbol{X}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{u}_i, \qquad 1.$$

where $\boldsymbol{X}_{ij}$ is the vector of covariates for the jth event of driver i, $\boldsymbol{X}_{ij} = (1, X_{1ij}, X_{2ij}, \ldots, X_{Kij})'$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)'$ is the vector of regression parameters; $\boldsymbol{u}_i$ is the driver-specific random effect; and $\boldsymbol{z}_{ij}$ is the corresponding row of design matrix.

In the context of an NDS, the $\boldsymbol{X}_{ij}$ represents potential risk factors. For example, $X_{1ij} = 1$ if the driver was using a cell phone at the jth event of driver i; $X_{1ij} = 0$ otherwise. The exponential of the corresponding regression coefficient $\exp(\beta_1)$ is the OR of cell phone use while driving. In the mixed effects logistic model, the observations from the same driver, i, share the same random term

u_i, which induces the driver-specific correlations. This mixed effect logistics regression has been the primary analysis model for the case-cohort study design (Dingus et al. 2016, Guo et al. 2017, Klauer et al. 2014). The model can be extended to a multilevel hierarchical model to accommodate multiple-level correlation structures (e.g., correlation among drivers from the same geographic regions in addition to the within-driver correlation).

Case-Crossover Design

The case-crossover design was originally proposed to study transient effects on the risk of acute events (Maclure 1991). The key characteristic of a case-crossover design is that the matched controls are selected from the same subject with the case, thus controlling for the subject-specific confounding effect. Redelmeier & Tibshirani (1997) used the case-crossover study design to evaluate the risk of cell phone use when driving, showing that cell phone use made the injury crash risk four times higher. McEvoy et al. (2005) conducted a case-crossover study design and reached similar conclusions. Note that the studies discussed above were based on field-collected crash data rather than NDS data.

The case-crossover design has been adopted to evaluate driver behavior in NDSs. Klauer et al. (2010) used the case-crossover method to evaluate the risk associated with eye-off driving tasks. Victor et al. (2015) used the case-crossover method to estimate the eye-off driving task and distraction risk using SHRP 2 NDS data. More recently, Owens et al. (2018) evaluated the crash risk associated with cell phone use while driving using the case-crossover approach. Guo et al. (2019) proposed a semiparametric Bayesian regression model for analyzing unbalanced case-crossover NDS data.

For each crash, the case-crossover approach samples matched controls from the same driver, and the controls are required to share the same values as a set of matching factors, e.g., time of day and road type. The controls in the case-crossover design are the same as control periods of driving from the case-cohort design. For example, in the case-crossover study by Guo et al. (2019) and Klauer et al. (2010), for each crash, attempts were made to identify 15 matched controls for each crash/near-crash by matching driver ID, day of week, time of day, and location (or road type). A recent American Automobile Association (AAA) Foundation for Traffic Safety study added additional matching factors such as speed, locality, relationship to junction, and traffic density (Owens et al. 2018). The matched controls were also required to have occurred prior to crashes in order to avoid any potential impacts of the crash on subsequent driver behavior. The case and controls within a matched stratum share the same values for all matching factors and are from the same driver, which can effectively control for the potential of confounding effects from the driver and matching factors.

While it is tempting to add more matching factors to control for as much confounding as possible, additional matching factors could also increase the difficulties in finding matched controls, especially for factors requiring a visual examination of the videos, such as traffic density and relationship to junction. Adding more factors could also lead to overmatching (Brookmeyer et al. 1986). When using this design, following matching principles is imperative—for example, requiring match factors to be associated both with crash risk and with risk factors to be evaluated.

The standard statistical model for the case-crossover study is conditional logistic regression. Conditional logistic regression eliminates strata effects by conditioning on $\sum_{j=1}^{n_i} Y_{ij} = 1$, i.e., each matched set contains one case and several controls:

$$\mathrm{P}\left(Y_{i1} = 1, Y_{i2}, \ldots\ Y_{i\mathrm{n_i}} = 0|X, \sum_{j=1}^{n_i} Y_{ij} = 1\right) = \frac{\exp\left(\beta X_{i1}\right)}{\sum_{j=}^{n_i} \exp\left(\beta X_{ij}\right)},$$

where Y_{ij} is the response variable ($1 = $ case, $0 = $ control) for the jth event in stratum i; X_{ij} is the indicator variable for a risk factor ($1 = $ exposed, $0 = $ unexposed); β is the regression coefficient; and the exponential of β, $\exp(\beta)$, represents the OR.

While a 1:4 case to control ratio is considered most efficient in a case-control study, Mittleman et al. (1995) suggest case-crossover can benefit from more matched samples (e.g., increasing the match ratio from 1:4 to 1:100 could lead to a 40% reduction in confidence interval). For NDS data, the challenge is that multiple matching factors often lead to situations in which no sufficient matched controls can be found. For example, Guo et al. (2019) attempted to find 1:15 matching, but only 40% of the cases found all 15 matched controls, which led to severely unbalanced data. Standard conditional logistic regression does not fit the unbalanced case-crossover well because the stratum effect, in which the stratum with more matched controls should contain more information, cannot be addressed by the standard conditional logistic regression model. Accordingly, Guo et al. (2019) proposed a semiparametric Bayesian model for the unbalanced case-crossover data.

Let Y_{ij} be a binary response for the jth response of the ith stratum. Let X_{ij} be the vector of covariates, for $j = 1, \ldots, m_i + 1$, and $i = 1, \ldots, N$, for a $1 : m_i$ matched design in which the number of matched controls m_i varies by stratum. The case status Y_{ij} is modeled as a function of the covariate X_{ij} in a prospective model form as follows. For simplification, a model with a single covariate ($p = 1$) is presented:

$$\mathrm{P}\left(Y_{ij} = 1 | X_{ij},\ \beta_{0ij}\right) = g\left(\beta_{0ij} + x_{ij}\beta\right),\ \beta_{0ij} = s_i + q_{ij},$$

where s_i is a stratum random effect and q_{ij} is a within-stratum random effect. This two-level hierarchical model catches both the within- and between-strata effects for unbalanced case-crossover data. In addition, the model addresses the issue that inference only depends on discordant pairs in regular conditional logistics regression models. Several alternatives have been proposed for the above model, including random effects and semiparametric Bayesian models, and study results indicate that the semiparametric Bayesian is preferred in terms of efficiency and robustness (Guo et al. 2019).

Alternative Reference Exposure Levels

To evaluate the OR of time-variant events, a reference exposure level is required. The reference level represents a base level a particular exposed level is compared with. This is particularly important for driver behavior research since a driver can engage in many types of risk behaviors, e.g., cell phone use and eating while driving. Whether other behaviors should be included in reference exposure level will affect the OR estimation considerably.

Two commonly used alternative reference levels are the model-driving and the all-driving levels. The model-driving reference level is when drivers are sober, alert, and attentive; that is, the driver does not show any signs of visually identifiable impairment, fatigue, or distraction (Dingus et al. 2016, Guo et al. 2017). This reference exposure level represents an ideal driver status and the OR thus represents the increase in risk caused by exposure to a factor (e.g., cell phone use while driving, relative to this ideal driver status). **Table 2** illustrates the alternative reference levels for cell phone distraction using the SHRP 2 NDS. The OR based on model-driving reference level is $\mathrm{OR1} = \frac{39^{*}9{,}279}{260^{*}627} = 2.22$.

Another reference exposure level is the all-driving or specific-task-absent level, which includes all events that are not exposed to the risk factor. In the example here, that would be cell phone use versus all other behaviors, i.e., the all-driving reference exposure level (Owens et al. 2018). The related OR from **Table 2** is $\mathrm{OR2} = \frac{39^{*}18{,}727}{808^{*}627} = 1.44$. The all-driving reference level represents a

Table 2 Reference exposure level comparison

	Exposed	Reference exposure level (model driving)	Other exposure	Reference exposure level (all driving)
	Handheld cell phone use	Sober, alert, and attentive (model driving)	Other distraction	Model driving and other distraction
Crashes	39	260	548	808
Control periods	627	9,279	9,448	18,727

mixture of driver behavior other than the one being evaluated. The OR should be interpreted as elevated risk associated with a specific behavior compared with a mix of all other driver behaviors.

The ORs based on the two alternative reference levels have different assumptions and different interpretations. The model-driving reference level offers an OR estimation compared with an ideal driving behavior. The all-driving reference indicates the risk compared with when the driver was not engaging in the specific behavior. As an all-driving reference level may include many other risky behaviors, it will typically lead to a lower OR compared with the model-driving-based OR.

One major disadvantage of the all-driving reference level is that it varies for different behaviors. For example, the all-driving reference level for cell phone use will be different from the reference level for eating while driving. This makes the interpretation and comparison of different driver behaviors more complicated. Furthermore, the all-driving reference level could change when drivers change their behavior; for example, drivers use cell phones much more than they did 10 years ago. Therefore, the OR based on the all-driving reference only represents the relative risk for a specific population in a specific time period. The model-driving level, in contrast, provides a consistent comparison level that is invariant for the risk factor to be evaluated and is stable over a long period of time.

One argument against the model-driving reference level is that it represents a status that a driver will not be likely to maintain for an extended period of time under real-life driving conditions. As different reference levels would lead directly to different ORs, caution should be used when selecting appropriate reference levels, and the corresponding ORs should be interpreted accordingly.

STATISTICAL MODELS FOR THE FREQUENCY OF CRASHES AND CRASH SURROGATES

Driving safety is, in general, evaluated by either crashes or crash surrogates (AASHTO 2010). The frequency of crashes or crash surrogates thus provides a direct measure of safety. Poisson and negative binomial (NB) regression models are the state-of-practice statistical models in traffic safety modeling and are the basis of the *Highway Safety Manual*'s Safety Performance Functions (AASHTO 2010). In the context of NDSs, the units of analysis are typically drivers or a prespecified driving period and the response is the frequency of crashes. The covariates are the factors associated with the driver during the specific driving period (e.g., driver age, gender, fitness-to-drive condition, personality, or frequency of kinematic events).

Poisson regression models are the base models for such analysis. The number of events is assumed to follow a Poisson distribution, or the related overdispersed NB distribution:

$$Y_i \sim \text{Poisson}\,(m_i\lambda_i) \;\text{ or }\; Y_i \sim \text{NB}(m_i\lambda_i, \gamma),$$

where Y_i is the number of events for the subject (driver or vehicle) i, λ_i is the expected event rate (i.e., the number of events per unit of exposure) for subject i, and the exposure m_i is usually measured by time or distance traveled. The NB distribution can accommodate overdispersion: $\mathrm{Var}(Y_i) = m_i\lambda_i + (m_i\lambda_i)^2\gamma$, where γ is the dispersion parameter. The expected rate of event λ_i is connected with potential risk factors via a link function. The logarithm link function is the most popular: $\log\lambda_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta}$, where $\boldsymbol{X}_{ij}$ is the vectors of potential risk factors, such as the age of a driver and fitness-to-drive condition.

The model setup described above has been widely used in NDS research. Fitch et al. (2013) used mixed effect Poisson regression to model the risk of a safety-critical event associated with cell phone use subtasks. Guo & Fang (2013) used NB regression to model the impact of different individual driver risk factors on crash and near-crash rates. Ouimet et al. (2014) used mixed effect longitudinal Poisson regression to measure the effect of cortisol hormones on teen drivers' driving risk over time. Chen et al. (2016) used the NB model to model the relationship between safety-critical event rates and truck drivers' sleep patterns adjusted by driver demographics. The NB regression model was also used to model the impacts of senior drivers' fitness-to-drive functional health conditions on their crash/near-crash risks, as well as the relationship between senior drivers' annualized mileage and their driving risk (Antin et al. 2017a,b; Guo et al. 2015).

Simons-Morton et al. (2011) used the following setup to model the rate of kinematic high G-force events, which occurred when the acceleration of the vehicle passed a threshold $\log\lambda_{ij} = \log m_{ij} + \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \boldsymbol{Z}'_{ij}b_i$, where m_{ij} is the exposure for driver i in time period j, b_i is a driver-specific random effect typically assumed to follow a normal distribution with zero mean and variance σ^2, $\boldsymbol{Z}'_{ij}$ is the vector of the design matrix for the random effect, and $\boldsymbol{X}_{ij}$ is row of the matrix of covariates.

The aforementioned mixed effect models rely on a relatively simple random effects structure. To accommodate multilevel correlation, Kim et al. (2013) proposed a hierarchical Bayesian model with the following form: $\log\lambda_{ij} = \log m_{ij} + \boldsymbol{X}'_{ij}\boldsymbol{\beta} + b_{ij}$, where $b_{ij} = b_i + O_{ij} + s_{ij}$, and b_i, O_{ij}, and s_{ij} reflect subject-specific, overdispersion, and time series correlation, respectively. The Ornstein–Uhlenbeck process was used to incorporate serial correlation for longitudinal data (Taylor et al. 1994). Variations of Poisson and NB models with different correlation structure will serve as the primary statistical models for an NDS when the response is event frequency.

RECURRENT EVENTS MODELS

Typically, a driver will have experienced multiple safety-critical events in an NDS. For example, the event time of the Naturalistic Teenage Driving Study is shown in **Figure 6**. The intervals between events provide valuable information that would be lost in an aggregated analysis.

Recurrent event models can incorporate the high-resolution NDS information, thus benefiting risk analysis by, for example, identifying a risk change-point. A commonly used recurrent event modeling framework is based on Poisson processes (Andersen & Gill 1982). The nonhomogeneous Poisson process (NHPP) is a Poisson process in which the intensity function varies across time. Let $N_j(t)$ be the number of events for driver j until time t. The $N_j(t)$ follows a Poisson distribution with rate parameter $\Lambda_j(t)$ as the cumulative intensity function. The derivative of $\Lambda_{\mathrm{j}}(t)$ is the intensity function $\lambda\mathrm{j}(t) = \frac{\mathrm{d}\Lambda\mathrm{j}(t)}{\mathrm{d}t}$.

Recurrent event models have been used to model a change in intensity over time, such as the risk change-point for novice teenage drivers, the impact of a crash on driving behavior, and the change in driver performance over long trips for commercial truck drivers (Chen & Guo 2016; Li et al. 2017, 2018). In the case of the first example, studies have shown that the driving risk of novice teenage drivers is highest during the initial period after licensure and is followed by a rapid decrease. The change-point in driving time provides an important measure of how much

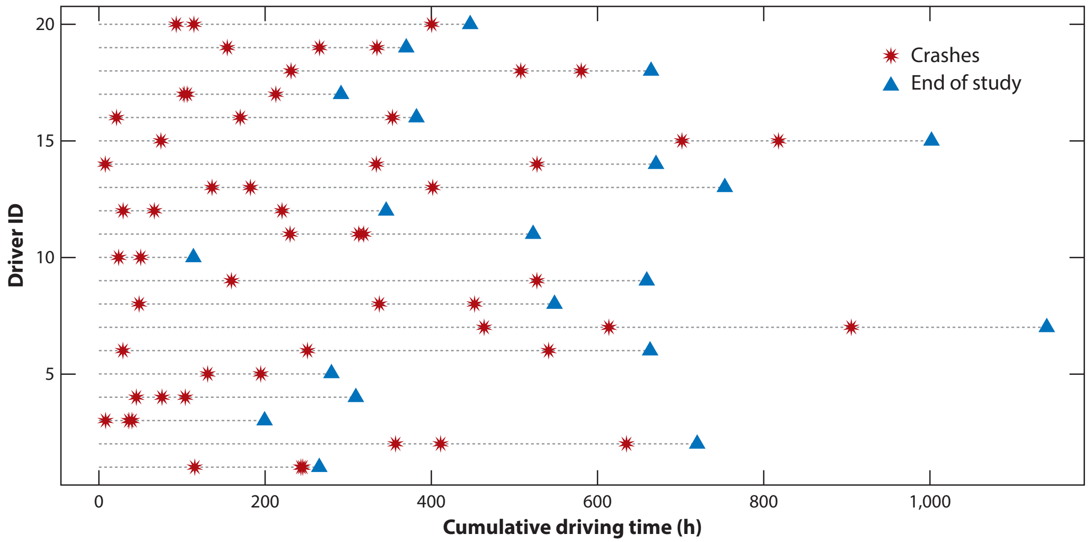

Figure 6

Event plot for a sample of 20 second Strategic Highway Research Program (SHRP 2) naturalistic driving study participants. Each line represents the driving time for a participant, asterisks represent crashes, and triangles represent the end of study for each driver.

experience is required before relatively safer driving occurs, a critical reference for graduated driver licensing regulation as well as teenage driver education programs.

Assuming identical change-points among drivers, Li et al. (2017) proposed using an identical-intensity model and a subject-specific intensity model to estimate the change-points by maximizing the profile likelihood using the Naturalistic Teenage Driving Study data. The number of change-points was selected using the Akaike information criterion. This method was an extension of the likelihood-based approach in Frobish & Ebrahimi (2009). Li et al. (2018) developed a hierarchical Bayesian finite mixture model to group drivers into clusters, with drivers within the same cluster sharing the same change-point and intensity functions. The number of clusters was selected using the deviance information criterion. The density of the change-point for subject j was assumed to be a finite mixture with K components: $f(\tau_j|\boldsymbol{\pi}_j, \boldsymbol{\mu}) = \sum_{k=1}^{K} \pi_{jk}\delta_{\mu_k}(\tau_j)$, where π_{jk}s were the mixture proportions, μ_k was the change-point value for the kth cluster, and $\delta_{\mu_k}(\tau_j)$ was a degenerate measure with probability 1 when $\tau_j = \mu_k$ and 0 otherwise. These studies indicated that the driving risk for novice teenage drivers decreased significantly after approximately 70 hours of independent driving. In addition, teenage driving risk varied substantially among individuals both in change-point and intensity rates, which is consistent with other studies (Guo et al. 2013).

An alternative recurrent event model is based on frailty models using random effects to accommodate heterogeneity across drivers. Random effects are typically assumed to have a multiplicative effect on intensity. A commonly used approach assumes a gamma frailty for the conditional Poisson process, which is $u_i\lambda_i(t)$, with $u_i \sim \mathrm{Ga}(1/\phi, 1/\phi)$, and yields a NB marginal process (Lawless 1987). Chen & Guo (2016) investigated the influence of crashes on driving risk using the 100-Car NDS data under the frailty framework. Four alternative recurrent event models were used to evaluate the influence of crashes based on actual driving time under the NHPP assumption.

Recurrent event models fit naturally with NDSs' event generation process. By utilizing high-resolution NDS data, recurrent event models may reveal the patterns of nonhomogeneous processes that are critical for traffic safety improvement measures.

DISCUSSION

NDSs contribute to traffic research by supplying large data sets of continuously collected, high-resolution, high-dimensional driving information. The rich and complex data provide unprecedented opportunity for understanding driving risk and driving behavior but can also present challenges in conducting statistical analyses. Findings from various NDSs have had profound impacts on public policy making, vehicle safety improvement, and driver education development.

As presented in this article, a wide variety of statistical and epidemiological methods can be used to work with NDS data. Regardless of the method ultimately selected, a key element to the success of research projects dealing with NDS data is assembling a research team with a domain expert, a statistician, and a data management/reduction specialist. The high costs that can be associated with NDS data analysis require a detailed data analysis plan to be put in place in the early stages of the project to ensure proper data reduction processes. A knowledgeable research team will ensure that this plan is created and carried out efficiently.

Current NDS data analyses primarily rely on epidemiological approaches and generalized linear regression models. With the public availability of the SHRP 2 NDS data set, and as subsequent NDS projects are completed, more diverse research questions can be addressed by NDS, for example, the safety of automatic driving. This will come with a need for alternative and/or additional statistical analysis methods. These might include recurrent event models, extreme value models, and/or causal inference models. Data mining and machine learning methods are also promising approaches for analysis of high-resolution, high-frequency NDS driving data.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENT

The author gratefully thanks an anonymous reviewer for the valuable comments and Anne Abramson and Douglas Beckner for their help with the article and figure editing.

LITERATURE CITED

AASHTO (Am. Assoc. State Highw. Transp. Off.). 2010. *Highway Safety Manual*. Washington, DC: AASHTO

Andersen PK, Gill RD. 1982. Cox's regression model for counting processes: a large sample study. *Ann. Stat.* 10:1100–20

Antin JF, Guo F, Fang YJ, Dingus TA, Hankey JM, Perez MA. 2017a. The influence of functional health on seniors' driving risk. *J. Transp. Health* 6:237–44

Antin JF, Guo F, Fang YJ, Dingus TA, Perez MA, Hankey JM. 2017b. A validation of the low mileage bias using naturalistic driving study data. *J. Safety Res.* 63:115–20

Behbahani. 2015. New time-based surrogate safety measure to assess crash risk in car-following scenarios. *Transp. Lett.* 7:229–38

Brookmeyer R, Liang K, Linet M. 1986. Matched case-control designs and overmatched analyses. *Am. J. Epidemiol.* 124:693–701

Chen C, Guo F. 2016. Evaluating the influence of crashes on driving risk using recurrent event models and Naturalistic Driving Study data. *J. Appl. Stat.* 43:2225–38

Chen GX, Fang YJ, Guo F, Hanowski RJ. 2016. The influence of daily sleep patterns of commercial truck drivers on driving performance. *Accid. Anal. Prev.* 91:55–63

Cook JK, Antin JF, Atkins WM, Hankey JM. 2015. *Naturalistic Driving Study: collecting data on cell phone use*. Rep. S2-S06-RW-2, Transp. Res. Board, Washington, DC

Cooper PJ. 1984. Experience with traffic conflicts in Canada with emphasis on "post encroachment time" techniques. In *International Calibration Study of Traffic Conflict Techniques*, ed. E Asmussen, pp. 75–96. Berlin: Springer

Dingus TA, Guo F, Lee S, Antin JF, Perez M, et al. 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *PNAS* 113:2636–41

Dingus TA, Hankey JM, Antin JF, Lee SE, Eichelberger L, et al. 2015. *Naturalistic Driving Study: technical coordination and quality control*. Rep. S2-S06-RW-1, Transp. Res. Board, Washington, DC

Dingus TA, Klauer SG, Neale VL, Petersen A, Lee SE, et al. 2006. *The 100-Car Naturalistic Driving Study: Phase II—results of the 100-Car field experiment*. Rep. DOT HS 810 593, NHTSA, US Dep. Transp., Washington, DC

Eenink R, Barnard Y, Baumann M, Augros X, Utesch F. 2014. *UDRIVE: The European naturalistic driving study*. Paper presented at Transport Research Arena 2014, Paris

Fitch GM, Soccolich SA, Guo F, McClafferty J, Fang Y, et al. 2013. *The Impact of Hand-held and Hands-free cell phone use on driving performance and safety-critical event risk*. Rep. DOT HS 811 757, NHTSA, US Dep. Transp., Washington, DC

Frobish D, Ebrahimi N. 2009. Parametric estimation of change-points for actual event data in recurrent events models. *Comput. Stat. Data Anal.* 52:671–82

Gettman D, Head L. 2003. Surrogate safety measures from traffic simulation models. *Transp. Res. Rec.* 1840:104–15

Glauz WD, Bauer K, Migletz DJ. 1985. Expected traffic conflict rates and their use in predicting accidents. *Transp. Res. Rec.* 1026:1–12

Guo F, Fang Y. 2013. Individual driver risk assessment using naturalistic driving data. *Accid. Anal. Prev.* 61:3–9

Guo F, Fang Y, Antin JF. 2015. Older driver fitness-to-drive evaluation using naturalistic driving data. *J. Safety Res.* 54:49.e29–54

Guo F, Hankey JM. 2009. *Modeling 100-Car safety events: a case-based approach for analyzing naturalistic driving data*. Rep., Natl. Surf. Transp. Saf. Cent. Excel., Va. Tech. Transp. Inst., Blacksburg, VA

Guo F, Kim I, Klauer SG. 2019. Semiparametric Bayesian models for evaluating time-variant driving risk factors using naturalistic driving data and case-crossover approach. *Stat. Med.* 38:160–74

Guo F, Klauer SG, Fang Y, Hankey JM, Antin JF, et al. 2017. The effects of age on crash risk associated with driver distraction. *Int. J. Epidemiol.* 46:258–65

Guo F, Klauer SG, Hankey JM, Dingus TA. 2010a. Near crashes as crash surrogate for naturalistic driving studies. *Transp. Res. Rec.* 2147:66–74

Guo F, Klauer SG, McGill MT, Dingus TA. 2010b. *Evaluating the relationship between near-crashes and crashes: Can near-crashes serve as a surrogate safety metric for crashes?* Rep. DOT HS 811 382, NHTSA, US Dep. Transp., Washington, DC

Guo F, Simons-Morton BG, Klauer SE, Ouimet MC, Dingus TA, Lee SE. 2013. Variability in crash and near-crash risk among novice teenage drivers: a naturalistic study. *J. Pediatr.* 163:1670–76

Hankey JM, Perez MA, McClafferty JA. 2016. *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Rep., Va. Tech. Transp. Inst., Blacksburg, VA

Hanowski RJ, Bowman D, Alden A, Wierwille WW, Carroll R. 2008. *PERCLOS+: Moving beyond single-metric drowsiness monitors*. SAE Tech. Pap. 0148–7191, Va. Tech. Transp. Inst., Blacksburg, VA

Heinrich HW, Petersen DC, Roos NR, Hazlett S. 1980. *Industrial Accident Prevention: A Safety Management Approach*. New York: McGraw-Hill

Hickman JS, Hanowski RJ, Bocanegra J. 2010. *Distraction in commercial trucks and buses: assessing prevalence and risk in conjunction with crashes and near-crashes*. Rep. FMCSA-RRR-10–049, Fed. Motor Carrier Saf. Adm., US Dep. Transp., Washington, DC

Jonasson JK, Rootzén H. 2014. Internal validation of near-crashes in naturalistic driving studies: a continuous and multivariate approach. *Accid. Anal. Prev.* 62:102–09

Kalra N, Paddock SM. 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transp. Res. A Policy Pract.* 94:182–93

Kim S, Chen Z, Zhang Z, Simons-Morton BG, Albert PS. 2013. Bayesian hierarchical Poisson regression models: an application to a driving study with kinematic events. *J. Am. Stat. Assoc.* 108:494–503

Klauer SG, Dingus TA, Neale VL, Sudweeks JD, Ramsey DJ. 2006. *The impact of driver inattention on near-crash/crash risk: an analysis using the 100-Car Naturalistic Driving Study Data*. Rep. DOT HS 810 594, NHTSA, US Dep. Transp., Washington, DC

Klauer SG, Guo F, Simons-Morton BG, Ouimet MC, Lee SE, Dingus TA. 2014. Distracted driving and risk of road crashes among novice and experienced drivers. *New Engl. J. Med.* 370:54–59

Klauer SG, Guo F, Sudweeks JD, Dingus TA. 2010. *An analysis of driver inattention using a case-crossover approach on 100-Car data*. Rep. DOT HS 811 334, NHTSA, US Dep. Transp., Washington, DC

Knipling RR. 2015. *Naturalistic driving events: no harm, no foul, no validity*. Paper presented at the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Salt Lake City, UT

Laureshyn A, Svensson Å, Hydén C. 2010. Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. *Accid. Anal. Prev.* 42:1637–46

Lawless JF. 1987. Negative binomial and mixed Poisson regression. *Can. J. Stat.* 15:209–25

Li Q, Guo F, Kim I, Klauer SG, Simons-Morton BG. 2018. A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers. *J. Appl. Stat.* 45:604–25

Li Q, Guo F, Klauer SG, Simons-Morton BG. 2017. Evaluation of risk change-point for novice teenage drivers. *Accid. Anal. Prev.* 108:139–46

Lu G, Liu M, Wang Y, Yu G. 2012. Quantifying the severity of traffic conflict by assuming moving elements as rectangles at intersection. *Procedia Soc. Behav. Sci.* 43:255–64

Maclure M. 1991. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.* 133:144–53

McEvoy SP, Stevenson MR, McCartt AT, Woodward M, Haworth C, et al. 2005. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 331:428–28

Minderhoud MM, Bovy PHL. 2001. Extended time-to-collision measures for road traffic safety assessment. *Accid. Anal. Prev.* 33:89–97

Mittleman MA, Maclure M, Robins JM. 1995. Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *Am. J. Epidemiol.* 142:91–98

NHTSA (Natl. Highw. Traffic Saf. Adm.). 2012. *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. Docket NHTSA-2010-0053, NHTSA, US Dep. Transp., Washington, DC

NHTSA (Natl. Highw. Traffic Saf. Adm.). 2013. *National motor vehicle crash causation survey: report to Congress*. Rep. DOT HS 811 059, NHTSA, US Dep. Transp., Washington, DC

NHTSA (Natl. Highw. Traffic Saf. Adm.). 2015. *Traffic safety facts: motor vehicle traffic crashes as a leading cause of death in the United States, 2010 and 2011*. Rep. DOT HS 812 203, NHTSA, US Dep. Transp., Washington, DC

NHTSA (Natl. Highw. Traffic Saf. Adm.). 2017. *Traffic safety facts 2015: a compilation of motor vehicle crash data from the Fatality Analysis Reporting System and the general estimates system*. Rep. DOT HS 812 384, NHTSA, US Dep. Transp., Washington, DC

Olson RL, Hanowski R, Hickman J, Bocanegra J. 2009. *Driver distraction in commercial vehicle operations*. Rep. FMCSA-RRR-09–042, Fed. Motor Carrier Saf. Adm., US Dep. Transp., Washington, DC

Ouimet MC, Brown TG, Guo F, Klauer SG, Simons-Morton BG, et al. 2014. Higher crash and near-crash rates in teenaged drivers with lower cortisol response an 18-month longitudinal, naturalistic study. *JAMA Pediatr*. 168:517–22

Owens JM, Dingus TA, Guo F, Fang Y, Perez M, McClafferty J. 2018. *Crash risk of cell phone use while driving: a case-crossover analysis of naturalistic driving data*. AAA Found. Traffic Saf., Washington, DC

Parker MR, Zegeer CV. 1989a. *Traffic conflict techniques for safety and operations—observers manual*. Rep. FHWA-IP-88–02, Fed. Highw. Adm., US Dep. Transp., Washington, DC

Parker MR, Zegeer CV. 1989b. *Traffic conflict techniques for safety and operations: Engineers guide*. Rep. FHWA/RD-84/042, Fed. Highw. Adm., US Dep. Transp., Washington, DC

Perkins SR, Harris JL. 1968. Traffic conflict characteristics-accident potential at intersections. *Highway Res. Rec.* 225:35–43

Prentice RL. 1986. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73:1–11

Redelmeier DA, Tibshirani RJ. 1997. Association between cellular-telephone calls and motor vehicle collisions. *New Engl. J. Med.* 336:453–58

Simons-Morton BG, Cheon K, Guo F, Albert P. 2013. Trajectories of kinematic risky driving among novice teenagers. *Accid. Anal. Prev.* 51:27–32

Simons-Morton BG, Ouimet MC, Zhang ZW, Klauer SE, Lee SE, et al. 2011. Crash and risky driving involvement among novice adolescent drivers and their parents. *Am. J. Public Health* 101:2362–67

Simons-Morton BG, Zhang Z, Jackson JC, Albert PS. 2012. Do elevated gravitational-force events while driving predict crashes and near crashes? *Am. J. Epidemiol.* 175:1075–79

Smadi O, Hawkins N, Hans Z, Bektas BA, Knickerbocker S, et al. 2015. *Naturalistic driving study: development of the Roadway Information Database*. Rep. S2-S04A-RW-1, Transp. Res. Board, Washington, DC

Smith BM, Dyer CR, Chitturi MV, Lee JD. 2017. Automatic driver head state estimation in challenging naturalistic driving videos. *Transp. Res. Rec.* 2663:48–56

Tarko AP. 2012. Use of crash surrogates and exceedance statistics to estimate road safety. *Accid. Anal. Prev.* 45:230–40

Taylor JMG, Cumberland WG, Sy JP. 1994. A stochastic model for analysis of longitudinal AIDS data. *J. Am. Stat. Assoc.* 89:727–36

Topp HH. 1996. *Traffic safety work with video processing*. Transp. Dep., Univ. Kaiserslautern

Treat JR, Tumbas N, McDonald S, Shinar D, Hume R, et al. 1979. *Tri-level study of the causes of traffic accidents: final report. Executive summary*. Rep., Inst. Res. Public Saf., Bloomington, IN

van Dongen HP, Jackson ML, Belenky G. 2010. *Duration restart period needed to recycle with optimal performance: phase II*. Rep. FMCSA-MC-RRR-10-062, Fed. Motor Carrier Saf. Adm., US Dep. Transp., Washington, DC

Victor T, Dozza M, Bärgman J, Boda C-N, Engström J, et al. 2015. *Analysis of naturalistic driving study data: safer glances, driver inattention, and crash risk*. Rep. 0309274230, Transp. Res. Board, Washington, DC

Vogel K. 2003. A comparison of headway and time to collision as safety indicators. *Accid. Anal. Prev.* 35:427–33

Wu K-F, Jovanis PP. 2013. Defining and screening crash surrogate events using naturalistic driving data. *Accid. Anal. Prev.* 61:10–22

Annual Review of Statistics and its Application

Volume 6, 2019

Contents

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics