Accident Analysis and Prevention 72 (2014) 210–218

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level

Kun-Feng Wu[a,*], Jonathan Aguero-Valverde[b], Paul P. Jovanis[c]

[a] Department of Transportation and Logistics Management, National Chiao Tung University, Taiwan
[b] Civil Engineering, University of Costa Rica, Costa Rica
[c] Civil and Environmental Engineering, Larson Institute, Penn State University, PA, United States

ARTICLE INFO

Article history:
Received 26 May 2012
Received in revised form 29 May 2014
Accepted 4 July 2014
Available online 31 July 2014

Keywords:
Naturalistic driving data
Surrogate events
Safety-related events
Multivariate Poisson log-normal models

ABSTRACT

There has been considerable research conducted over the last 40 years using traffic safety-related events to support road safety analyses. Dating back to traffic conflict studies from the 1960s these observational studies of driver behavior have been criticized due to: poor quality data; lack of available and useful exposure measures linked to the observations; the incomparability of self-reported safety-related events; and, the difficulty in assessing culpability for safety-related events. This study seeks to explore the relationships between driver characteristics and traffic safety-related events, and between traffic safety-related events and crash involvement while mitigating some of those limitations. The Virginia Tech Transportation Institute 100-Car Naturalistic Driving Study dataset, in which the participants' vehicles were instrumented with various cameras and sensors during the study period, was used for this study. The study data set includes 90 drivers observed for 12–13 months driving. This study focuses on single vehicle run-off-road safety-related events only, including 14 crashes and 182 safety-related events (30 near crashes, and 152 crash-relevant incidents). Among the findings are: (1) drivers under age 25 are significantly more likely to be involved in safety-related events and crashes; and (2) significantly positive correlations exist between crashes, near crashes, and crash-relevant incidents. Although there is still much to learn about the factors affecting the positive correlation between safety-related events and crashes, a Bayesian multivariate Poisson log-normal model is shown to be useful to quantify the associations between safety-related events and crash risk while controlling for driver characteristics.

1. Introduction

Road crashes are frequently characterized as rare events, often requiring many years of observation before an underlying mean can be reliably estimated. As an alternative, researchers have studied a range of safety-related events, which are similar to crashes in terms of crash risk, but without an impact (e.g. Perkins and Harris, 1967; Evans and Wasielewski, 1982; Hydén, 1987). Although driver characteristics have been shown to be associated with crash risk (e.g. Shinar, 2007; Dewar and Olson, 2001), the relationships between these characteristics and traffic safety-related events, and between safety-related events and crash involvement have been challenging to assess. Data have been of variable quality across studies because accuracy depends on the training of individual human observers. Exposure variables such as miles-traveled are typically not available for the drivers under observation, so crash rates cannot be computed. Many methods require the use of self-reported crashes and safety-related events, which raises issues concerning reliability. Finally, the culpability of the driver for the safety-related event may be difficult to determine. The methodological objective of this study is to utilize naturalistic driving study data with state-of-the-art statistical models to gain insights about driver behavior, crashes and safety related events. It is hoped that the model framework illustrates the utility of the method, so other users of naturalistic driving data may confidently adopt similar approaches.

1.1. Safety-related events analysis

There has been considerable research conducted over the last 45 years concerning the development of safety-related events for assessing traffic safety (e.g. Perkins and Harris, 1967; Datta, 1979; Hauer, 1982; Evans and Wasielewski, 1982, 1983; Risser, 1985; Hydén, 1987; Chin and Quek, 1997; Shankar et al., 2008; Tarko et al.,

* Corresponding author. Tel.: +886 3 571 6440.
E-mail addresses: kfwu@nctu.edu.tw (K.-F. Wu), jonathan.aguero@ucr.ac.cr (J. Aguero-Valverde), ppj2@psu.edu (P.P. Jovanis).

http://dx.doi.org/10.1016/j.aap.2014.07.005
0001-4575/

2009; Jovanis et al., 2010; McGehee et al., 2010; Guo et al., 2010; Wu and Jovanis, 2012, 2013). In the past, the goal of safety-related events research was driven by the perceived need to conduct safety analyses (e.g. identification of hazardous sites or evaluation of the effectiveness of safety countermeasure) more quickly (before a large number of crashes occur) and with more data than are typically available from law-enforcement-reported crash records (Datta, 1979; Grayson and Hakkert, 1987; Archer, 2004). Human factors have been generally considered to be among the most important factors in crash occurrence, and therefore, the relationship between driver characteristics, driving behavior, and crash involvement has also been studied using safety-related events for decades (e.g. Evans and Wasielewski, 1982, 1983; Wagenaar and Reason, 1990; Verschuur and Hurts, 2008). These types of research are often referred to as analysis of safety-related events, near crashes, risky driving, near misses, or surrogate events. The distinction between safety-related and surrogate events are that surrogate events include both crashes and near crashes with common etiologies to crashes, whereas safety-related events include only near crashes, risky driving, or near misses (Wu and Jovanis, 2012, 2013). The relationship between traffic safety-related events and crash frequency is typically studied at either segment/intersection or driver level as multiple events could occur at the same intersection/segment, and the same driver may encounter multiple events during a period of time (Jovanis et al., 2012).

The most well-known and studied safety-related event is the traffic conflict occurring at intersections. In one of the first conflict studies (Perkins and Harris, 1967), conflicts were defined based on evasive actions taken by drivers such as the appearance of brake lights or sudden lane changes. The general approach of a traffic conflict study is to collect crash and conflict data from a number of intersections, and estimate the “conversion factor,” connecting the number of conflicts and traffic volume to the number of crashes (Hydén, 1987; Wu and Jovanis, 2012). Sayed and Zein (1999) conducted a similar study to validate the relationship, and they reported a statistically significant relationship between three-year crash frequency and observed conflicts at signalized intersections. Recently researchers have started utilizing field operational test or naturalistic driving study data to study run-off-road (ROR) safety-related event (e.g. Leblanc et al., 2006; Hallmark et al., 2011; Wu and Jovanis, 2012; Gordon et al., 2013).

The idea that a crash is preceded by factors more remote in time and place from the crash has been proposed for some time (e.g. Evans and Wasielewski, 1982, 1983; Evans, 1991; Dewar and Olson, 2001; Wagenaar and Reason, 1990). These factors include driver characteristics and driving behavior. Hence, safety-related driver behavior models have been proposed to connect crash involvement, driver characteristics, and driving behaviors (e.g. Verschuur and Hurts, 2008). Some of the earliest driver-based studies were conducted by Evans and Wasielewski (1982, 1983). Their studies measured following headways on the roads, and photographically recorded the license plate number to obtain the information regarding the vehicle and the vehicle owner, including age, gender, and driving record. It was shown that drivers with prior crash involvement and traffic violations were more likely to be observed at such risky headways; whereas seat-belted drivers tend to avoid risky headways. In addition, drivers under age 30, tend to take more risk in everyday driving in terms of short headways. Risser (1985) found that the sum of all errors in driving behavior shows correlation with the subjects’ accidents in the past five years as well as with the subjects’ traffic conflicts during a 1-h driving test.

For many years, safety-related events studies have been conducted by collecting field data at intersection/segments or by interviewing drivers. For research at intersection/segment level, researchers in the 1980s video-taped vehicle maneuvers at intersections (e.g. Hydén, 1987; Evans and Wasielewski, 1982). In recent years, researchers have started using high-definition street cameras and image recognition techniques to streamline data collection and analyses (e.g. Chin and Quek, 1997). For research at the driver level, a typical approach is to query a sample of drivers with a driver behavior questionnaire (DBQ), including self-reported driving exposure, errors, and traffic law violations, and then associate the DBQ with the frequency of their crash involvement in the past (e.g. Verschuur and Hurts, 2008).

1.2. Data analysis issues and naturalistic driving study data

Safety-related event analyses have been limited by the need to obtain exposure measures such as miles-traveled (e.g. Evans and Wasielewski, 1983). Without considering driving exposure, one would expect the safety-related events occur more frequently with tasks and activities that drivers perform more frequently (Hanowski et al., 2005). Second, self-reported safety-related events or behavior are subject to a variety of biases, making them subjective and difficult to compare across studies. Third, researchers have stressed the need to distinguish the culpability of the driver for risky driving events and crashes, i.e. to focus on those events for which the drivers are at fault (af Wåhlber, 2003). These limitations are mostly due to the constraint of data collection. In addition, traffic safety-related events are often defined using a single measure such as time to lane departure (e.g. UMTRI; Hallmark et al., 2012). Recent research has shown that there is a need to define safety-related events while accounting for event attributes (e.g. whether driver was distracted) and driving environment (e.g. daytime or nighttime condition) (Davis and Hourdos, 2012; Wu and Jovanis, 2012). Event attributes and driving environment necessitate advanced data collection techniques.

Naturalistic driving studies provide an opportunity to more precisely observe and measure safety-related events (e.g. Bareket et al., 2003; Dingus et al., 2005). Stutts et al. (2005) installed unobtrusive video units in the vehicles of 70 volunteer drivers over one-week time period to study drivers’ exposure to distractions. Hanowski et al. (2005) also collected real-world driving data from truck drivers, and found that a small number of long-haul drivers were involved in a disproportionate number of distraction-related safety-related events. Guo et al. (2010) utilized the Virginia Tech Transportation Institute (VTTI) 100-Car Naturalistic Driving Study dataset to show the association between crashes and near crashes. All of these studies concluded that naturalistic driving studies could provide a useful supplement to more controlled laboratory and field studies to further our understanding of the effects of driver characteristics on traffic safety. Naturalistic driving studies cannot only precisely measure driving exposure, but also more plausibly identify the culpability of risky driving events and crashes, and disentangle different types of crashes and different causes.

There are two primary distinguishing features for a naturalistic driving study (Jovanis et al., 2011):

1. Vehicles are instrumented with video camera technologies that record the driver and the road ahead of the vehicle continuously during driving. In addition to the video, other on-board sensors continuously record vehicle accelerations in three dimensions and well as rotational motion along the same axes. Radars are often present to record proximity to other vehicles and potential obstacles on the roadway or roadside. All these data are recorded and stored within an on-board data acquisition system (i.e. DAS).
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.