The association between crashes and safety-critical events: synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers

Miao Cai[a], Mohammad Ali Alamdar Yazdi[b], Amir Mehdizadeh[c], Qiong Hu[c], Alexander Vinel[c], Karen Davis[d], Fadel Megahed[e], Hong Xian[a], Steven E. Rigdon[a,*]

[a] *Department of Epidemiology and Biostatistics, Saint Louis University, Saint Louis, MO, 63103, United States*
[b] *Carey Business School, Johns Hopkins University, Baltimore, MD, 21218, United States*
[c] *Department of Industrial and Systems Engineering, Auburn University, Auburn, AL, 36849, United States*
[d] *Department of Computer Science and Software Engineering, Miami University, Oxford, OH, 45056, United States*
[e] *Department of Information Systems and Analytics, Miami University, Oxford, OH, 45056, United States*

Abstract

The past decade has witnessed continuous growth of naturalistic driving studies (NDSs). In NDSs, safety-critical events (SCEs) are commonly used to measure safety since crashes are very rare. However, the association between SCEs and crashes is not consistent in previous studies and has not been verified among commercial truck drivers. Based on routinely collected kinematic data from 31,828 truck drivers in a large commercial trucking company, this paper examines the association between four types of SCEs (headways, hard brakes, collision mitigation, and rolling stability) and crashes, as well as injuries and fatalities. Bayesian negative binomial models were applied to examine this association among drivers in different business units and driver types, as well as all drivers combined. It was found that a unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% credible interval CI: 8-8.8%) increase in crashes per mile and 8.7% (95% CI: 4.8-13.6%) increase in the number of injuries per mile. The increase was different in different types of SCEs: 3.3% (95% CI: 2.6-4%) for headways, 8.1% (95% CI: 7.5-8.7%) for hard brakes, 50.4% (95% CI: 41.4-60%) for rolling stability, and 22.2% (95% CI: 19.8-24.5%) for collision mitigation. The results are consistent when stratified by different business units and driver types. This study provides statistically strong and robust evidence that SCEs are positively associated with crashes and injuries among commercial truck drivers. NDS and kinematic data routinely collected by trucking or insurance companies provide a promising opportunity for future data analytics research.

Keywords: truck, naturalistic driving studies, safety-critical events, crashes, injuries, fatalities

1. Introduction

According to the World Health Organization (WHO), road injuries were the eighth leading cause of death globally in 2018 (The WHO, 2018a). Globally, road injuries claim around 1.35 million lives annually, which accounts for over 2.2% of the annual deaths worldwide (The WHO, 2018a). Among children and young adults aged five to 29 years old, road traffic injuries are the leading cause of death (The WHO, 2018b). Besides life losses, road injuries also cause up to 50 million non-fatal injuries and around 75.5 million disability-adjusted life years (Staton et al., 2016, The WHO, 2018b). Among different types of vehicles on the road, trucks are often associated with catastrophic consequences, which can be attributable to long routes, intensive schedule, massive size and weight, and potentially

*Corresponding Author

Email addresses: miao.cai@slu.edu (Miao Cai), yazdi@jhu.edu (Mohammad Ali Alamdar Yazdi), azm0127@auburn.edu (Amir Mehdizadeh), qzh0011@auburn.edu (Qiong Hu), alexander.vinel@auburn.edu (Alexander Vinel), davisk4@miamioh.edu (Karen Davis), fmegahed@miamioh.edu (Fadel Megahed), hong.xian@slu.edu (Hong Xian), steve.rigdon@slu.edu (Steven E. Rigdon)

hazardous cargoes (Hanowski et al., 2005, Chen and Xie, 2014, Chen et al., 2016, Zheng et al., 2018). For this reason, truck crash prediction has been a widely studied topic in traffic safety research.

A substantial amount of road crash studies focused on using driver characteristics, traffic, weather, and others to predict the likelihood, frequency, or severity of crashes based on retrospective police report data (Lord and Mannering, 2010, Wang et al., 2013, Savolainen et al., 2011, Wang et al., 2013, Roshandel et al., 2015). However, crash data used in these studies ultimately trace back to retrospective reports, and therefore are subject to several limitations (Stern et al., 2019). First, crash data reported by drivers and police are post-hoc and therefore subject to recall and information bias. The reports cannot accurately record the information including real-time traffic and weather variables prior to the accidents. Second, crash data, especially those associated with trucks, are extremely sparse compared to non-crashes. The Federal Motor Carrier Safety Administration (2018) estimated that fatalities associated with large trucks and buses in 2017 were 0.156 per million traveled miles. Because this is so low, it is difficult to make statistically valid conclusions on the risk factors of crashes (Guo et al., 2010, Theofilatos et al., 2018). Third, drivers at fault tend to underreport or not report critical factors such as distraction or cell phone use, in order to avoid associated penalties (Dingus et al., 2011, Stern et al., 2019). Finally, those crashes with no human injuries and little financial losses also tend to be underreported or not reported.

In view of the limitations in traditional crash data studies, naturalistic driving studies (NDSs) have been proposed to collect high-resolution objective data and enhance traffic safety (Mehdizadeh et al., 2020). NDSs use sensors and devices to proactively and unobtrusively collect high-frequency kinematic and Global Positioning System (GPS) data in a real-world driving setting (Neale et al., 2005, Guo and Fang, 2013, Guo, 2019). Compared to traditional truck crash studies that are roadway-based or intersection-based case-control studies, NDS data are vehicle-based and are superior in estimating the rate or risk of events since all non-events are collected. More importantly, NDS has the power of providing precise real-time data on kinematic events, such as acceleration or deceleration based on a pre-specified threshold, which provide the opportunity to investigate the short period prior to crashes or safety events without information bias or reporting bias (Guo et al., 2010). NDS data are typically collected at a high frequency;; for example, the acceleration and video data in the second Strategic Highway Research Program (SHRP2) NDS (Hankey et al., 2016) and 100-Car NDS (Dingus et al., 2006) were collected continuously at a rate of 10 hertz from ignition-on to ignition-off. Other studies, especially those involving GPS tracking collect data less frequently, say once every 10 seconds to 10 minutes. Thus, the amount of NDS data are generally very large, which provides both an opportunity and a challenge for data analytics.

The number of NDSs worldwide has been increasing in the past 10 years. For example, the 100-Car NDS (Dingus et al., 2006), SHRP2 (Guo, 2019), and the Europe's UDRIVE NDS (Eenink et al., 2014). These NDSs are sponsored by government departments or research organizations, and the number of participating drivers are typically limited considering the budget of the programs. However, in the recent years, it is noticeable that an increasing number of vehicle and insurance companies have been routinely collecting real-time NDS and kinematic data, which have been applied in safety surveillance, insurance pricing, and performance evaluation fields. If these routinely collected large-scale high-resolution data are sufficiently valuable to give insights into transportation safety, they hold huge potential for follow-up data analytics research.

This study is done in collaboration with a leading freight shipping truck company in the United States. The name of the company cannot be revealed for confidentiality reasons and we will refer to it as Company A hereafter. The company provides transportation, delivery, and logistics services to customers and consumers in North America. We consider three business units (dedicated, intermodal, and final-mile) and three driver types (local, regional, and over-the-road). Drivers in the dedicated business unit serve a single customer with familiar routes, task and work duties. Intermodal drivers work with major rail providers and transport freight container from rail yards to customer locations, which are traditional driving duties. Final-mile drivers receive tasks from their managers every day and make deliveries to customer locations. Regarding driver types, local drivers transport freight within a 200-mile radius and return home the same day, regional drivers move freights in a region that may include surrounding states, and over-the-road drivers pick up and delivery freights throughout the country, and they are required to be on duty for at least two weeks and then rest.

The primary objective of this paper is to investigate the association between crashes and safety-critical events (SCEs) among commercial truck drivers. Detected from dynamic kinematic events, SCEs are special types of accident precursors that have all features of accidents, except that potentially catastrophic outcomes were avoided by last-second evasive maneuvers (Dingus et al., 2011, Saleh et al., 2013). These SCEs are around 10 to 15 times as frequent as crashes, and past studies suggested that they are indicative of crashes (Guo et al., 2010). SCEs provide an alternative measure of transportation safety, without having to observe for a prohibitively long time to have a sufficient number of crashes (Guo, 2019). However, previous studies that quantify this association are limited in the number of drivers and no studies have examined this relationship among commercial truck drivers.

This paper sets the foundation for future studies modeling SCEs using data analytical methods by examining the association between SCEs and crashes, as well as injuries and fatalities. The purpose of this study is to explain this association. The question of what factors affect the likelihood of a crash, or predict crashes (Shmueli et al., 2010) is a separate topic which we do not address here. We address the following questions:

(A) To what extent are SCEs associated with crashes, injuries, and fatalities among commercial truck drivers?

(B) Is this association consistent for all four different SCEs (headways, hard brakes, collision mitigation, and rolling stability) ?

(C) Does this association hold in different types of business units (dedicated, intermodal, and final-mile) and driver types (local, regional, and over-the-road)?

By answering the questions above, this paper contributes to the following aspects. First, NDS data from more than 30,000 commercial truck drivers were collected, and the data cover most areas in the United States and are generally representative of the trucking industry. The large sample size provides a large statistical power to detect potential relationship between SCEs and crashes. Second, four types of SCEs are considered in this study: headways, hard brakes, collision mitigation, and rolling stability. We show that the magnitude of association between crashes and the four types of SCEs are different. Third, since most of the crashes have no human injuries or fatalities, sample sizes in previous studies are ususally too small to detect the influence of predictor variables. As the number of drivers and miles driven is relatively large, the association between injuries, fatalities, and SCEs is examined, which is a more important issue than crashes since most crashes have no human injuries.

2. Literature review

There has been an increasing number of studies since the 1980s evaluating the association between crashes and surrogate measures, with the general approach being to estimate the conversion factor between the two types of events (Evans and Wasielewski, 1982, 1983, Cooper, 1984, Risser, 1985, Hydén, 1987). This topic has become a crucial issue as NDS data sets are becoming increasingly available to researchers in the recent decade. The first large-scale NDS was the 100-Car NDS, including 100 drivers in northern Virginia and Washington, D.C., in the United States (Neale et al., 2005). The study continuously and naturalistically recorded driving data for 102 participating drivers for one year, resulting in around two million vehicle miles and over 40,000 hours of driving data. Over the study period, 69 crashes, 761 near-crashes, and 8,295 crash-relevant conflicts and proximity conflicts were collected. Based on this 100-Car study, Dingus et al. (2006) found that hard braking events were significantly associated with collisions and near-crashes. Since the number of near-crashes and incidents were significantly larger than crashes, they proposed using near-crashes and incidents as surrogate measures of crashes.

The rationale for using near-crashes and SCEs as surrogates for crashes is the Heinrich's Triangle, which assumes that less severe events are more frequent than severe events, and the frequency of severe events can diminish as the frequency of less severe events decreases (Guo, 2019). The former assumption is commonly seen in most NDSs and is considered to be reasonable (Guo et al., 2010). The association in the latter assumption can be quantitatively tested based on crash and SCEs data in large-scale NDSs, although the causal mechanism is not clear and hard to prove (Guo et al., 2010). The validity of SCEs varies significantly from studies to studies, so researchers are advised to be cautious when applying SCEs as the outcome (Guo, 2019). If the assumptions of Heinrich's Triangle are met, SCEs can provide insights to crash risk and transportation safety when the number of crashes is limited. However, the crucial question prior to using SCEs as a measure of real crashes is whether they are good surrogates of traffic crashes.

Guo et al. (2010) proposed two critical principles for using near-crashes as surrogates for crashes: 1) similar or the same causal mechanisms between crashes, surrogates, and risk factors, 2) a strong association between the frequency of surrogates and crashes. Based on the 100-Car study, they investigated the two principles using a sequential factor analysis, a Poisson regression, and sensitivity analyses. The study concluded that using near-crashes as surrogates for crashes will lead to conservative risk estimates but significantly reduce the variance of estimation. They suggested

Table 1: Differences between our study and existing literature regarding sample size, driver type, crash surrogates, region, time frame, and statistical models.

Authors	sample size	Driver type	Crash surrogates	Country or region	Time frame	Statistical models
Dingus 2006	100 cars	general	braking, steering, accelerating	Northern Virginia & Washington, DC, USA	1 year	95% confidence limits modeled as a Poisson distribution
Guo 2010	100 cars	general	near crashes	Northern Virginia & Washington, DC, USA	1 year	sequential factor analysis, Poisson regression
Gorden 2011	78 drivers	general	lateral deviation, lane-departure warning, time to edge crossing	Southeastern Michigan, USA	10 months	seemingly unrelated regression, Poisson regression
Simons-Morton 2012	42 drivers	newly licensed teenagers	elevated gravitational-force	Virginia	18 months	logistic regression using generalized estimating equations
Wu 2012	100 cars	general	lateral and longitudinal acceleration, Event button, Yaw rate, Forward amd rear time-to-collision	Northern Virginia & Washington, DC, USA	1 year	logistic regression
Guo 2013	100 cars	general	critical-incident events	Northern Virginia & Washington, DC, USA	1 year	negative binomial regression
Wu 2014	100 cars	general	safety-related events	Northern Virginia & Washington, DC, USA	1 year	Poisson regression
Pande 2017	33 drivers	general	high magnitude jerks while decelerating	San Luis Obispo, California, USA	10 days	negative binomial regression
Gitelman 2018	64 vehicles	general	braking, speed alerts	Israel	1 year	negative binomial regression
This paper	**31,828 drivers**	**commercial truck drivers**	**headways, hard brakes, collision mitigation, rolling stability**	**National data, USA**	**1 year**	**Bayesian negative binomial regression**

that using near-crashes as surrogates in small-scale studies will be informative for evaluating the risk of crashes.

Gordon et al. (2011) conducted a preliminary study to validate surrogates for road-departure crashes by spatially merging road geometry, average traffic, crashes, and NDS data. Bayesian seemingly unrelated Poisson models estimated with weighted least squares were used to examine whether these predictor variables have the same effects on crashes and surrogates. They found that time to edge crossing and lane-departure warning were two useful surrogates for crashes on rural nonfreeway roads, but lane deviation was a poor surrogate for lane-departure crashes.

Simons-Morton et al. (2012) examined whether elevated gravitational-force predicts crashes and near-crashes among 42 newly licensed teenage drivers in Virginia. The study used the Naturalistic Teenage Driving Study that followed the recruited drivers for 18 months. A logistic regression estimated with generalized estimating equations to account for the within-subject correlation among different months was applied. The study found that the rate of elevated gravitational-force events was positively associated with the rate of crashes and near-crashes (odds ratio: 1.07, 95% confidence interval: 1.02-1.12), with an area under the curve (AUC) value of 0.76.

Wu and Jovanis (2012) proposed a conceptual framework to estimate the crash-to-surrogate ratio π and used the 100-Car study to test the framework. The study found that the conditional probability of a crash was increased by a factor of 24 with a lateral acceleration more than 0.7 g, but the probability was decreased by other factors such as the event occurring in daylight and dry pavement. A later study by Wu and Jovanis (2013) developed diagnostic procedures to screen crashes and near misses under NDS settings. The study applied the 100-Car NDS on the proposed framework and identified three conditions to define surrogate events: 1) a maximum lateral acceleration difference of no smaller than 0.4 g, 2) non-intersection related, and 3) maximum lateral acceleration difference no smaller than 0.9 per event or between 0.8 and 0.9 g during night time.

Guo and Fang (2013) attempted to identify risk factors of driving at driver's level and predict high-risk drivers based on the 100-Car study. The study used a negative binomial regression to examine the potential four risk factors of crashes and near-crashes. They used a K-mean clustering algorithm to classify the drivers into high-, moderate-, and low-risk groups based on crash and near-crash rates, and applied two logistic regressions to predict high- or moderate-risk drivers. The results confirmed that critical-incident event rates were significantly associated with individual driving risk. The two logistic regressions achieved AUC values of 0.938 and 0.93. They also highlighted that it was a first-step study and more studies with larger and representative data were needed to confirm the association. A similar study by Wu et al. (2014) also used the 100-Car NDS data set. This study used a Bayesian multivariate Poisson log-normal model to simultaneously account for crash frequency and severity. They also found a significant positive association between crashes, near-crashes, and crash-relevant incidents.

Pande et al. (2017) used linear referencing to link Global Positioning System (GPS) data with roadway features on 39 segments of Highway 101 in California. Negative binomial models and random-effects NB models that account for segment-specific variance were used to investigate the relationship between historic crashes and hard braking events. It was found that the freeway segments with high hard braking rates also had higher long-term crash rates, although the other three explanatory variables, average daily traffic, the presence of horizontal curvature, and auxiliary lanes were not statistically significant.

Gitelman et al. (2018) used data collected by in-vehicle data recorders on 3,500 segments of interurban roads in Israel to examine the association between two types of safety-related events (braking and speed alert) and crashes on different road types. NB models were used to account for the over-dispersion in the data, and the covariates included several road infrastructure characteristics. The number of braking events was found to be positively associated with injury crashes on single- and dual-carriageway roads, while the association was not significant on freeways. Counterintuitively, they found that speed alert events were consistently and negatively associated with injury crashes on all road types. They suggested that speed alert events were not a good surrogate for crashes, possibly due to its rough definition.

However, some researchers have been skeptical of NDS. For example, Knipling (2015) challenged the validity of using NDS data and SCEs by arguing that the purpose of traffic safety studies is to identify causes of crash harm, where crash harm is defined as property damage, injury, income lost, and all other consequences of different severities (Zaloshnja and Miller, 2007). However, NDS often use SCEs as surrogates of crashes, but very few or no crashes are involved, let alone human harm. Therefore, it is argued that SCEs are not an appropriate part of the Heinrich's Triangle and researchers generally cannot derive valid quantitative conclusions on causations of harm based on NDS data sets. Another study by Knipling (2017) specifically targeted Hour-of-Service (HOS) rule research, such as Blanco et al. (2011) and Hanowski et al. (2008), and relevant policy revisions among commercial truck drivers. He argued that HOS studies with a quasi-experiment design were subject to confounding variables, so these studies are limited in demonstrating a causal relationship between HOS and safety outcomes. The paper also argued that NDS lacked external validity since no large truck NDS had examined the causal link between crashes and SCEs. Lastly, the construct validity was doubted since the relationship between driver fatigue, HOS, and SCEs had not been validated.

In summary, no consistent conclusion has been reached on the association between the number of crashes and surrogates in NDS (Guo, 2019). As shown in Table 1 and the above review, there are three gaps in previous NDS studies on the association between crashes and SCEs: 1) the number of drivers involved are limited to between 33 and 100, which makes it difficult to reach statistically significant conclusions on crashes, injuries, and fatalities, 2) although two separate NDS data sets : the Drowsy Driver Warning System Field Operation Test and the Naturalistic Truck Driving Study have been sponsored by the Federal Motor Carrier Safety Administration (Hickman et al., 2018), no studies have examined the association between crashes and surrogates using NDS data sets that specifically target commercial truck drivers, 3) since previous NDS data sets are limited in the number of drivers and mileage driven, no studies have examined the association between crashes surrogates and human injuries or fatalities.

3. Methodology

3.1. Data description

The data used in this study were initially collected by Company A for routine performance monitoring and driver assistance, not for research purpose as was the case in the 100-Car study. From April 1, 2015 to March 31, 2016, Company A installed data acquisition systems to their trucks in all three business units. These systems intermittently collected real-time driving "ping" data every couple of seconds to around 15 minutes: over 50% of the time intervals between two pings were less than 5 minutes and over 95% of them were less than 15 minutes. The time intervals varied among drivers, places, and trips, and there were no clear pattern of the intervals. Each ping is a data point that includes the exact date and time of the record (year, month, day, hour, minute, and second), GPS (latitudes and longitudes), GPS quality, speed, and drivers' anonymized unique ID. The research team excluded 2,520 (7.4%) drivers who 1) had less than 100 pings, 2) cannot be matched in driver demographics table, and 3) were identified as obvious outliers regarding the rates of SCEs. As shown in Table 3, a total of $1,494,678,173$ pings collected from 31,828 commercial truck drivers were included in the current study, and 98.7% of these pings were

Table 2: A description of the data sources

Ping		SCEs		Crashes		Trips	
Total pings	1,494,678,173	Total SCEs	450,758	Crashes	34,884	Total trips	18,740,142.00
Drivers	31,828	Headway	170,421	Injuries	239	Total miles	2,320,967,467.00
		Hard brakes	218,419	Fatalities	22	Total hours	65,646,731.00
		Collision mitigation	55,243			Median miles	77.06
		Rolling stability	6,675			Median hours	2.61

marked as having good quality (expected accuracy of 0.1-1.0 meters). The latitudes and longitudes had at least five decimal places, which were worth up to 1.1 meters, indicating high-precision GPS location data. The geographical distributions of these included pings are shown in the Results Section below.

Collected independently from the ping data, $450,758$ SCEs were recorded when pre-specified kinematic thresholds were triggered while driving. The collected SCE data include the exact date and time, latitude and longitude (specific to five decimal places), driver, and type of SCEs. This SCEs data included $170,421$ headways, $218,419$ hard brakes, $55,243$ collision mitigation, and $6,675$ rolling stability. Non-kinematic definitions of the four types of SCEs are:

- *Headway*: the distance between a lead vehicle and the truck is less than a preset threshold value. 118 SEC on the Excel book,
- *Hard brake*: the speed decrease within a unit time is larger than a preset threshold value. 9.5 MPHPS, but how to interpret it
- *Collision mitigation*: an imminent collision with a forward slower moving or standing vehicle or object. No threshold value on the Excel book.
- *Rolling stability*: potential roll-over or loss-of-control situations in various road or weather conditions (dry, wet, snow and ice-covered pavement). No threshold value on the Excel book.

A crash data set that included driver ID, state, city, report time, the number fatalities, and the number of injuries, were collected by retrospective reports from Company A. In total, 34,884 crashes, 239 injuries, and 22 fatalities were collected for included drivers in this table. The company also provided a driver demographics table that includes age, gender, race, business units, and driver types associated with the driver ID. All data were de-identified prior to the research team's access. The study protocol was approved by the Institutional Review Board of Saint Louis University.

3.2. Data aggregation

Since the original ping data set is gigantic and difficult to interpret regarding the routing, we aggregated them to trips using the approach adopted by Pande et al. (2017). A trip is defined as a continuous period of driving with no more than 30 minutes' stop. In practice, we sorted the original ping data according to date and time for each driver. Then, if the ping speed data showed that the truck was not moving (the speed of the ping equals zero) for more than 30 minutes, the ping data were separated into two different trips.

This data aggregation process is demonstrated in Figure 1, where the x-axis shows the date and time of pings, and the y-axis presents the speed (in miles per hour, MPH). Each point represented a ping at that date and time,

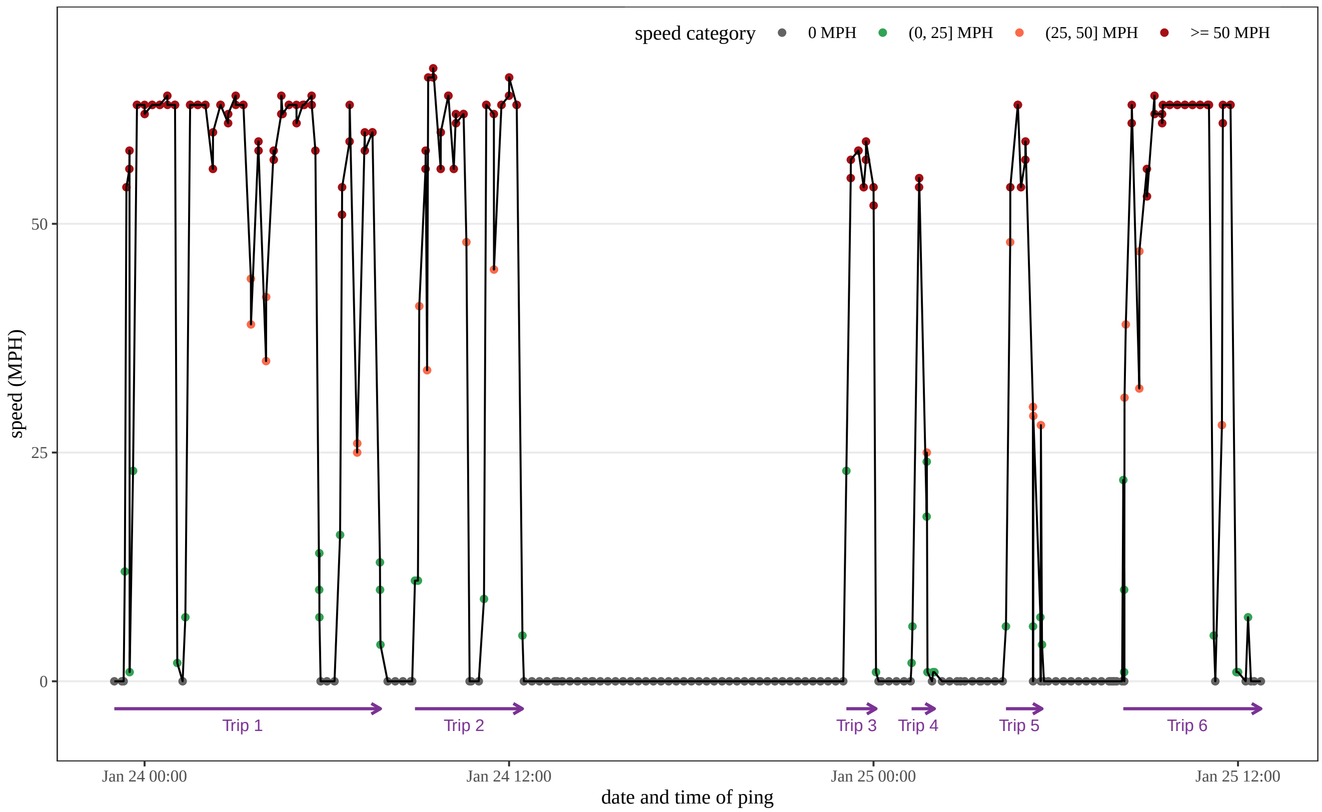


Figure 1: Data aggregation from ping data to trips for a sample of a commercial NDS data set.

with different colors indicating the real-time speed category. Whenever the truck stopped (the grey points) for at least 30 minutes, the pings were separated into different trips, indicated by the purple arrows in the bottom (Trip 1, Trip 2, . . ., Trip 6). The trip time was then calculated by taking the difference between the trip end time and start time.

The traveled distance within a trip was approximately calculated in three steps: 1) sort the ping data according to driver, date, and time, 2) compute the distance between each two nearest ping locations for each driver using the haversine method, which assumes a spherical earth and ignores ellipsoidal effects (Sinnott, 1984), 3) sum up all the distance traveled within a trip for each driver. This algorithm aggregated the original ping data into $18,740,142$ trips, with a total of $2,320,967,467$ miles traveled in $65,646,731$ hours. The median distance of a trip is 2.61 hours and the median number of miles per trip was 77.06 miles.

3.3. Statistical models

Since the outcome variable (the number of crashes for each driver) is a count variable, which is a strictly non-negative integer, Bayesian negative binomial models are used in this study. Compared with Poisson model commonly used for modeling count outcomes, negative binomial models can adjust for the variance independently from its mean, which handles potential overdispersion or underdispersion issues in the data. Let Y_i denote outcome variable (the number of crashes, injuries, and fatalities respectively) over a distance of T_i miles for the ith driver.

Each of the three outcomes were modeled in different Bayesian negative binomial models. We assume that Y_i has a negative binomial distribution with the mean parameter μ_i and a common auxiliary parameter ϕ. The corresponding probability mass function of Y_i is parameterized as

$$P(y_i|\mu, \phi) = \binom{y_i + \phi - 1}{y_i} \left(\frac{\mu}{\mu + \phi}\right)^{y_i} \left(\frac{\phi}{\mu + \phi}\right)^{\phi}, y_i = 0, 1, 2, \ldots. \tag{1}$$

The mean and variance of Y_i are $E[Y_i] = \mu$ and $V(Y_i) = \mu + \frac{\mu^2}{\phi}$. The inverse of ϕ controls the overdispersion, which is scaled by μ^2. By assuming that the number of SCEs per 10,000 miles has a mathematically multiplicative effect on the logarithm of rate of crashes μ_i, we have the following log-linear Bayesian negative binomial regression

$$\begin{aligned} Y_i &\sim \text{Negative Binomial}(T_i \times \mu_i,\ \phi) \\ \log \mu_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}, \end{aligned} \tag{2}$$

where β_0 is the intercept, $\beta_k,\ k = 1, 2, \cdots, K$ is the coefficient of the k-th predictor variable x_{ik}. The value of K is slightly different in non-stratified models ($K = 12$ for four SCEs model and $K = 9$ for other models) and stratified models by business units and driver types ($K = 8$). We take relatively noninformative priors for β_k and ϕ. Specifically, we assume $\beta_0, \beta_1, \cdots, \beta_K \sim \text{Normal}(0, 10^2)$ and $\phi \sim \text{Exponential}(1)$. The total miles driven T_i is considered as an offset variable to account for the mileage difference among the included drivers. Based on previous studies in the literature review section and data availability, the predictor variables of interest are the rates of all SCEs (the number of SCEs per 10,000 miles) and the rates of four different types of SCEs, including headways, hard brakes, collision mitigation, and rolling stability. The covariates were average speed, driver's age and gender, business units, and driver types. Business units included dedicated (reference group), intermodal, and final-mile. Driver types included local (reference group), regional, and over-the-road.

We applied Hamiltonian Monte Carlo to approximate the posterior distributions of all parameters. To make sure the Markov chain Monte Carlo converged, we set 4,000 iterations for each of the four chains, with the first 2,000 being warm-up iterations. The Markov chains were considered as converged when the Gelman-Rubin diagnostic $\hat{R}$ was less than 1.1 for each variable (Gelman et al., 1992). The posterior mean and 95% credible intervals (CIs) of the incidence rate ratios ($\exp(\beta)$) were reported. The interpretation of the incidence rate ratios in this Bayesian negative binomial model is: as the number of SCEs per 10,000 miles increases by one unit, the number of crashes per mile will be multiplied by $\exp(\beta)$ times. A 95% credible interval is the posterior probability that the parameter of interest falls into that range given the data is 95%. If the 95% CI of the incidence rate ratio includes one, then one is plausible value for the true incidence rate ratio, so we will consider the parameter of this variable as statistically insignificant. On the other hand, if the 95% CI of the incidence rate ratio excludes one, we will consider the parameter as statistically significant.

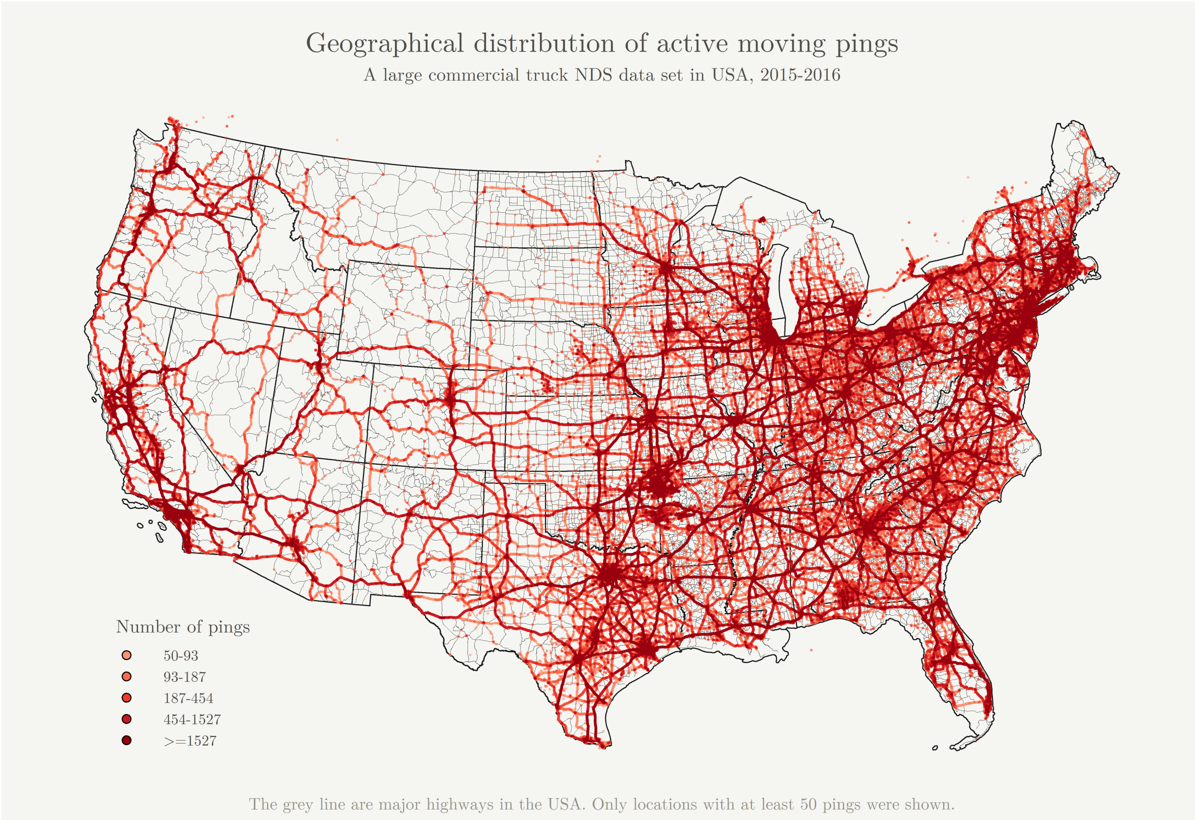
We used Pareto smoothed importance-sampling leave-one-out (PSIS-LOO) cross-validation to check the goodness-of-fit of and compare different models (Vehtari et al., 2015, 2017). Instead of exact cross-validation that refits the model with different subsamples, the PSIS-LOO uses fast, efficient, and stable importance sampling weights to approximate leave-one-out cross-validation (Gelfand et al., 1992, Gelfand, 1996). It estimates the expected log predicted density (ELPD), estimate number of parameters, and the LOO Information Criterion (LOOIC) for a new data set. Compared with other statistics such as Widely Applicable Information Criterion, Deviance information criterion, and other variants (Spiegelhalter et al., 2002, Watanabe, 2010), PSIS-LOO is both fast and stable in computing. Apart from PSIS-LOO, we also used posterior predictive checks to examine the prediction accuracy (Gelman et al., 2013, Chapter 6). The interpretation of these goodness-of-fit and model comparison statistics will be explained in the Results section. We used correlation plots and variance inflation factor tests to detect potential issue of multicollinearity. A Pearson correlation coefficient of more than 0.8 or a variance inflation factor score higher than 10 was indicative of multicollinearity in the models.

All data cleaning, statistical modeling, and visualization were performed in the statistical computing environment `R 3.6.0` (R Core Team, 2019). The haversine method distance was computed using the `distHaversine()` function in `geosphere` package (Hijmans, 2019). The Bayesian negative binomial model was conducted using the `stan_glm()` function in `rstanarm` package (Goodrich et al., 2018). The PSIS-LOO statistics were computed using the `loo()` function in `loo` package (Vehtari et al., 2019). Based on the Findable, Accessible, Interoperable, Reusable (FAIR) principles, a sample of the original ping data, full data for Bayesian negative binomial regression models, as well as an RMarkdown file documenting R code for data aggregation, Bayesian negative binomial regression models, model comparison, and diagnostic statistics are provided in the supplementary materials.

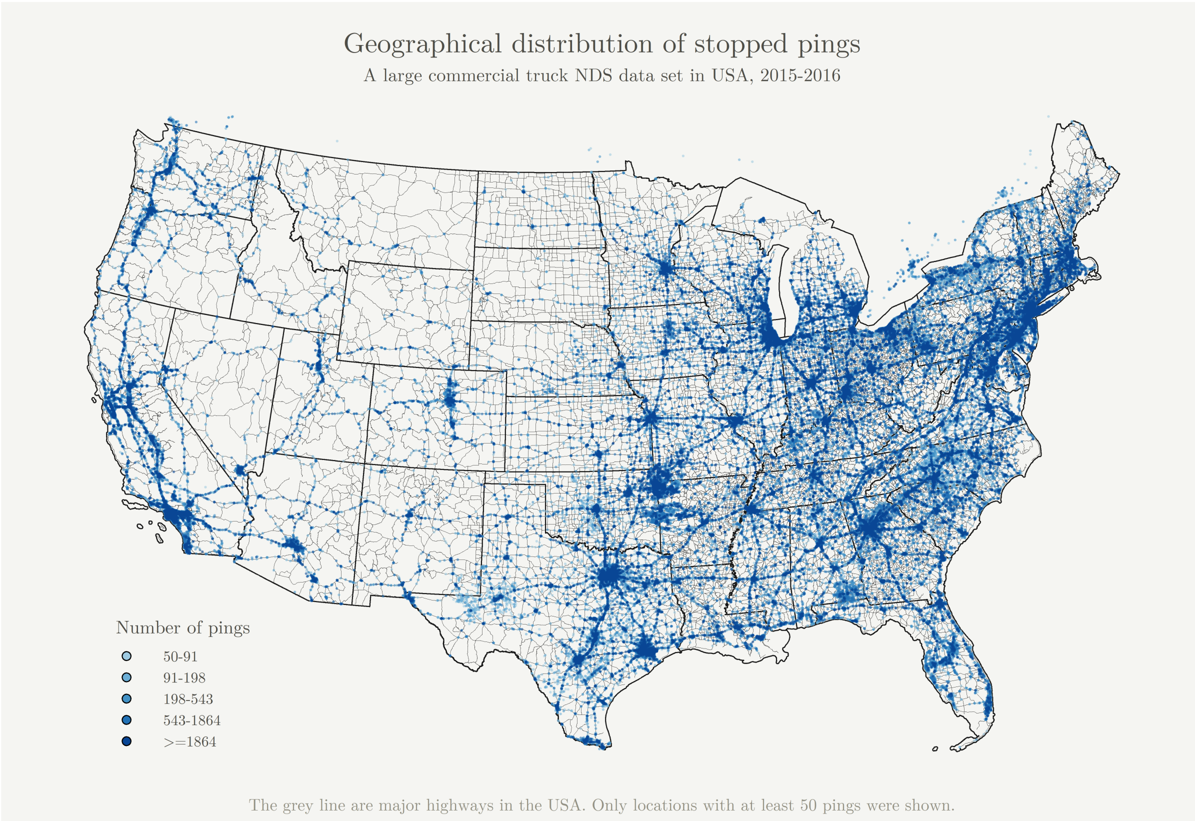
4. Results

4.1. Sample description

Figure 2 below shows the geographical point patterns of active pings that have a speed of more than 0 MPH (Figure 2a) and inactive pings that have a speed of 0 MPH (Figure 2b). Since plotting the original gigantic ping data may overwhelm the entire map, we rounded all latitudes and longitudes to two decimal places, which are approximately as large as normal-size parking lot, and then displayed only locations with more than 50 pings. The red points are the locations with at least 50 actively moving pings and blue dots are the locations with at least 50 stopping pings, with darker colors (dark red and dark blue) indicating higher number of pings. The background grey lines are the major roads in the United States (United States Geological Survey, 2014) and the solid black lines are state boundaries. The geographical point patterns suggest that most of the trucking transportation is closely matching population density (i.e., is more concentrated along the coasts). The active and inactive pings are generally consistent, but active pings are much more concentrated than inactive pings on the major Midwest major roads.



(a) Active pings



(b) Inactive pings

Figure 2: Geographical point patterns of active and inactive pings in a large commercial NDS data set.

Table 3: A summary of driver characteristics and predictor variables

Variable	Statistics
Rate of headways (mean (SD))	5.35 (27.01)
Rate of hard brakes (mean (SD))	6.86 (22.75)
Rate of rolling stability (mean (SD))	0.21 (1.20)
Rate of collision mitigation (mean (SD))	1.74 (4.66)
Age (mean (SD))	44.48 (11.72)
Ping speed (mean (SD))	29.85 (7.23)
Gender (%)	
Female	1,583 (5.0)
Male	29,248 (91.9)
Unknown	997 (3.1)
Business unit (%)	
Dedicated	16,152 (50.7)
Final-mile	5,908 (18.6)
Intermodal	9,768 (30.7)
Driver type (%)	
Local	13,381 (42.0)
Over-the-road	2,740 (8.6)
Regional	15,707 (49.3)

According to Table 3, among all the included 31,828 drivers, 29,248 (91.9%) were male, 1,583 (5.0%) were female, and 997 (3.1%) were unspecified. The mean age of the drivers were 44.48 years with a standard deviation of 11.72 years. The range of the ages were from 20 to 82 years. Regarding business units, 16,152 (50.7%) drivers served in the dedicated unit, in which the drivers work for a single customer and drive on familiar routes. 9,768 (30.7%) drivers were in inter-modal, where the drivers conduct traditional truck driving duties and transport freight containers from rail yards to local locations. 5,908 (18.6%) drivers were in final-mile and made deliveries to customer locations. As for driver types, 13,381 (42%) were local drivers who transport freight within 200 miles and return home on the same day. 15,707 (49.3%) were regional drivers who move freight within a region or interstate and return home on a weekly or bi-weekly basis. Only 2,740 (8.6%) were over-the-road drivers who deliver and pick up freight throughout the country, and they are required to be on the road for at least two weeks.

4.2. Bayesian negative binomial models for all included drivers

Table 5 below shows the estimates of posterior incidence rate ratios and their CIs in the Bayesian negative binomial models for all the included drivers. We ran six models with various predictor variables. These involved:

1. All SCEs combined (Pooled model),
2. Four types of SCEs as four variables in one model (Four SCEs),
3. Headways only,
4. Hard breaks only,
5. Rolling stability only,
6. Collision mitigation only.

Need clarification on models results. In the Bayesian setting, parameters are considered as random variables that have probabilistic distributions instead of unknown fixed values, so no P-values are reported here. Instead, we

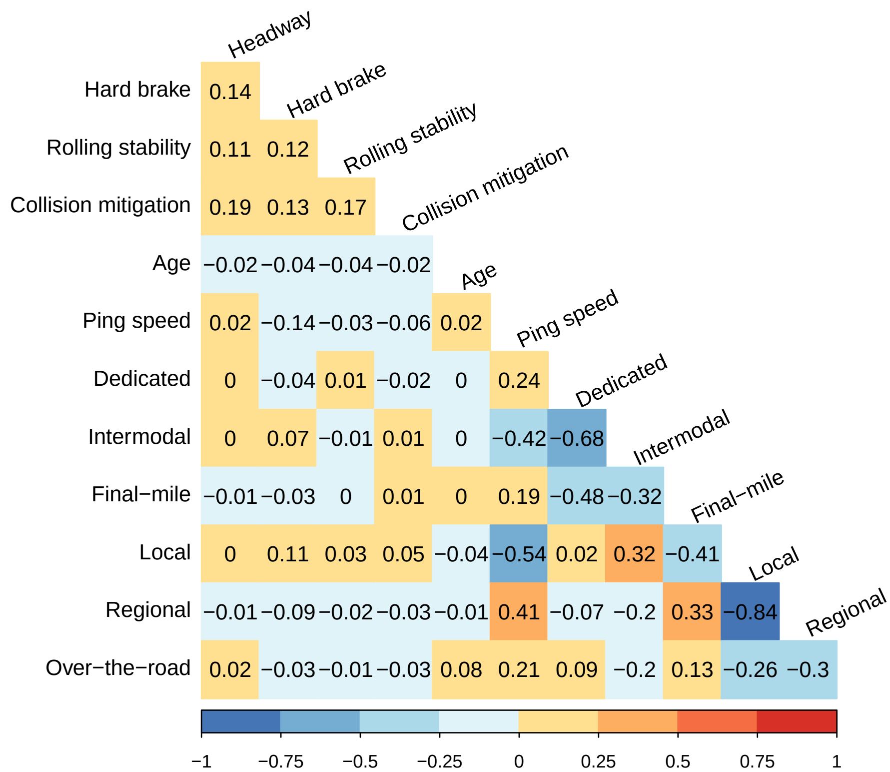


Figure 3: Correlation plot of the predictor variables.

reported 95% posterior credible intervals (CIs), which can be interpreted as there is 0.95 probability that these random parameters are within the fixed CIs given the observed data. All the incidence rate ratios of SCEs in the six models were greater than one and none of the associated CIs included one, indicating statistically strong evidence that the rates of SCEs were positively associated with the rates of crashes. The pooled model suggests that one unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% CI: 8-8.8%) increase in the number of crashes per mile. Specifically, four SCEs model (column 3 in Table 5) suggests that as the number of headway, hard brakes, rolling stability, and collision mitigation per 10,000 miles increased by one unit, the number of crashes per mile were increased by 3.3% (95% CI: 2.6-4%), 8.1% (95% CI: 7.5-8.7%), 50.4% (95% CI: 41.4-60%), and 22.2% (95% CI: 19.8-24.5%) respectively, controlling for other covariates. Figure 3 suggests that the correlation between the rate of four SCEs were minor (< 0.2). The variance inflation factor test in Table also indicate no multicollinearity in the models since all test statistics were less than 2 (Fox and Monette, 1992).

Table 4: Variance inflation factor test for multicollinearity.

Variables	Headways	Hard brakes	Rolling stability	Collision mitigation	Age	Ping speed	Gender	Business unit	Driver type
$\text{GVIF}^{(1/(2\text{DF}))}$	1.030	1.031	1.024	1.038	1.006	1.270	1.005	1.096	1.140

4.3. Bayesian negative binomial models stratified by business units and driver types

Need clarification on models results. Since the driving behaviors and routing patterns vary significantly among different business units and driver types, we also conducted Bayesian negative binomial models (four SCEs) for drivers in different business units and types, and the results are shown below in Table 6. The posterior incidence rate ratios and CIs of four SCEs are consistent with those in Table 5. All four type of SCEs were positively associated with the number of crashes per mile. None of the CIs included one except for headways in the dedicated & over-the-road unit. These stratified results indicate strong evidence that SCEs were positively associated with crashes in different business units and driver types. Among the four types of SCEs, rolling stability had the highest incidence rate ratios, followed by collision mitigation, hard brake, and headway.

Table 5: Bayesian negative binomial regressions with the rate of SCEs predicting crashes, non-stratified models

	Models					
Model names	Pooled model	Four SCEs	Headways	Hard brakes	Rolling stability	Collision mitigation
(Intercept)	0.054 (0.047, 0.062)	0.048 (0.042, 0.054)	0.090 (0.079, 0.103)	0.057 (0.050, 0.066)	0.082 (0.072, 0.093)	0.073 (0.064, 0.083)
All SCEs	1.084 (1.080, 1.088)					
Headways		1.033 (1.026, 1.040)	1.077 (1.069, 1.085)			
Hard brakes		1.081 (1.075, 1.087)		1.109 (1.102, 1.116)		
Rolling stability		1.504 (1.414, 1.600)			2.147 (2.015, 2.295)	
Collision mitigation		1.222 (1.198, 1.245)				1.343 (1.316, 1.369)
Age	0.992 (0.990, 0.993)	0.992 (0.991, 0.993)	0.989 (0.988, 0.990)	0.991 (0.989, 0.992)	0.989 (0.988, 0.991)	0.990 (0.988, 0.991)
Mean speed	0.979 (0.976, 0.982)	0.982 (0.979, 0.985)	0.971 (0.968, 0.973)	0.980 (0.977, 0.983)	0.973 (0.970, 0.976)	0.975 (0.973, 0.978)
Gender: male	0.817 (0.756, 0.886)	0.808 (0.754, 0.867)	0.848 (0.785, 0.919)	0.823 (0.762, 0.887)	0.845 (0.787, 0.909)	0.826 (0.770, 0.891)
Gender: unknown	0.975 (0.785, 1.199)	0.954 (0.777, 1.149)	1.097 (0.896, 1.347)	1.096 (0.884, 1.349)	1.018 (0.842, 1.239)	1.058 (0.870, 1.299)
Business unit: Inter-modal	0.698 (0.670, 0.727)	0.717 (0.690, 0.745)	0.706 (0.679, 0.735)	0.701 (0.672, 0.730)	0.735 (0.706, 0.765)	0.729 (0.700, 0.758)
Business unit: Final-mile	0.907 (0.861, 0.954)	0.897 (0.852, 0.943)	0.925 (0.882, 0.971)	0.904 (0.865, 0.948)	0.922 (0.880, 0.967)	0.901 (0.859, 0.942)
Driver type: over-the-road	1.071 (0.994, 1.151)	1.094 (1.022, 1.174)	1.053 (0.981, 1.131)	1.064 (0.994, 1.140)	1.067 (0.990, 1.144)	1.106 (1.030, 1.182)
Driver type: Regional	1.003 (0.957, 1.045)	1.012 (0.969, 1.057)	0.971 (0.928, 1.015)	0.994 (0.950, 1.037)	0.973 (0.932, 1.016)	0.984 (0.943, 1.028)
Fit statistics:						
sample size	31828	31828	31828	31828	31828	31828
elpd_loo	−39985.2 (236.5)	−39770.2 (233.5)	−40792.7 (238.9)	−40315.5 (237.2)	−40710.1 (237.8)	−40503.2 (239.4)
p_loo	18.1 (1.1)	30 (2.4)	19.8 (1.9)	18.2 (1.2)	15.9 (0.8)	16.1 (1)
looic	79970.4 (472.9)	79540.5 (467.1)	81585.4 (477.8)	80631 (474.5)	81420.1 (475.7)	81006.5 (478.7)

The SCEs were measured as the number of events per 10,000 miles driven. Incidence rate ratios and associated 95% credible intervals were reported.

Table 6: Bayesian negative binomial regressions with the rate of SCEs predicting crashes, stratified by business units and driver types

	Business units and driver types						
Business unit:	Dedicated	Dedicated	Dedicated	Inter-modal	Inter-modal	Final-mile	Final-mile
Driver type:	Local	Over-the-road	Regional	Local	Regional	Over-the-road	Regional
(Intercept)	0.055 (0.040, 0.076)	0.015 (0.008, 0.027)	0.062 (0.046, 0.084)	0.026 (0.020, 0.033)	0.021 (0.013, 0.033)	0.047 (0.021, 0.102)	0.033 (0.022, 0.049)
Headways	1.026 (1.011, 1.042)	1.001 (0.993, 1.010)	1.048 (1.032, 1.067)	1.026 (1.012, 1.042)	1.060 (1.038, 1.082)	1.082 (1.020, 1.149)	1.050 (1.031, 1.068)
Hard brakes	1.069 (1.057, 1.080)	1.241 (1.194, 1.293)	1.163 (1.140, 1.188)	1.047 (1.040, 1.054)	1.114 (1.093, 1.138)	1.086 (1.049, 1.131)	1.183 (1.154, 1.211)
Rolling stability	1.528 (1.367, 1.733)	1.648 (1.269, 2.229)	1.676 (1.467, 1.951)	1.419 (1.284, 1.578)	2.477 (1.590, 3.717)	4.320 (2.210, 9.522)	1.175 (1.039, 1.369)
Collision mitigation	1.163 (1.127, 1.203)	1.318 (1.132, 1.540)	1.362 (1.292, 1.440)	1.212 (1.174, 1.252)	1.577 (1.422, 1.766)	1.134 (0.952, 1.353)	1.170 (1.121, 1.234)
Age	0.992 (0.989, 0.995)	0.988 (0.982, 0.993)	0.993 (0.990, 0.996)	0.995 (0.993, 0.998)	0.986 (0.982, 0.990)	0.999 (0.989, 1.010)	0.997 (0.993, 1.000)
Mean speed	0.976 (0.970, 0.983)	1.016 (1.005, 1.027)	0.968 (0.962, 0.974)	0.994 (0.987, 1.000)	1.000 (0.988, 1.012)	0.973 (0.958, 0.988)	0.983 (0.973, 0.994)
Gender: male	0.883 (0.702, 1.083)	0.868 (0.631, 1.227)	0.844 (0.716, 0.997)	0.749 (0.650, 0.862)	0.841 (0.691, 1.029)	0.675 (0.433, 1.027)	0.751 (0.634, 0.893)
Gender: unknown	1.065 (0.589, 1.908)	1.378 (0.706, 2.617)	0.576 (0.325, 0.980)	1.287 (0.774, 2.079)	0.194 (0.044, 0.626)		0.816 (0.571, 1.158)
Fit statistics:							
sample size	6950	1797	7405	6429	3339	943	4963
elpd_loo	−9300.8 (125.3)	−2416.9 (52.2)	−9799.1 (112.1)	−7624.2 (90.8)	−3912.6 (70.5)	−1139.8 (40)	−5293.9 (85.4)
p_loo	30.6 (5)	14.3 (2.5)	20.4 (2.6)	17.4 (2.3)	13.9 (1.7)	11 (1.6)	19.4 (2.9)
looic	18601.6 (250.6)	4833.8 (104.5)	19598.1 (224.2)	15248.3 (181.6)	7825.2 (141)	2279.5 (79.9)	10587.7 (170.8)

The SCEs were measured as the number of events per 10,000 miles driven. Incidence rate ratio and associated 95% credible intervals were reported.

4.4. Bayesian negative binomial models for injuries and fatalities

Table 7 below presents the results of Bayesian negative binomial models predicting the number of injuries and the number of fatalities respectively. Compared with the models for crashes, the result for injuries and fatalities are less conclusive since the number of injuries and fatalities is small. One unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.7% (95% CI: 4.8%-13.6%) increase in the number of injuries per mile. When stratified into four different types of SCEs, all 95% CIs of incidence rate ratios included one, which indicated weak evidence for modeling injuries or fatalities, although the posterior means were positive. In the two models using the number of fatalities as the outcome variable (column 4 and 5), all 95% CIs of incidence rate ratios included one and the CIs were very wide, which suggested insufficient sample size to yield statistically significant results.

Table 7: Bayesian negative binomial regressions with the rate of SCEs predicting injuries and fatalities

Outcome variables	Injuries: pooled	Injuries: four SCEs	Fatalities: pooled	Fatalities: four SCEs
Intercept	0.013 (0.002, 0.070)	0.012 (0.002, 0.064)	0.008 (0.000, 1.855)	0.011 (0.000, 5.179)
All SCEs	1.087 (1.048, 1.136)		0.973 (0.791, 1.149)	
Headways		1.061 (0.961, 1.181)		0.955 (0.592, 1.478)
Hard brakes		1.080 (0.995, 1.177)		0.957 (0.652, 1.387)
Rolling stability		1.773 (0.684, 5.439)		1.631 (0.043, 102.782)
Collision mitigation		1.174 (0.987, 1.535)		0.866 (0.200, 3.632)
Age	0.987 (0.970, 1.004)	0.986 (0.969, 1.004)	0.966 (0.912, 1.020)	0.965 (0.906, 1.030)
Mean speed	0.967 (0.929, 1.007)	0.970 (0.931, 1.009)	0.915 (0.797, 1.049)	0.910 (0.778, 1.050)
Gender: male	0.825 (0.301, 2.149)	0.800 (0.298, 2.176)	1.770 (0.074, 54.444)	1.953 (0.062, 80.045)
Gender: unknown	1.022 (0.094, 8.499)	0.993 (0.092, 9.338)	0.093 (0.000, 76.248)	0.093 (0.000, 115.564)
Business unit: Inter-modal	0.459 (0.265, 0.788)	0.467 (0.280, 0.789)	0.354 (0.068, 1.573)	0.341 (0.044, 2.057)
Business unit: Final-mile	0.710 (0.352, 1.420)	0.675 (0.330, 1.321)	1.576 (0.209, 10.438)	1.536 (0.140, 13.475)
Type: Over-the-road	0.785 (0.321, 1.942)	0.801 (0.306, 1.955)	0.410 (0.022, 5.402)	0.388 (0.014, 6.205)
Type: Regional	0.472 (0.265, 0.821)	0.463 (0.263, 0.820)	0.389 (0.064, 1.970)	0.379 (0.050, 2.214)
Fit statistics:				
sample size	31828	31828	31828	31828
elpd_loo	−1134.5 (80.8)	−1137.3 (81.1)	−182.4 (37.9)	−182.4 (37.9)
p_loo	13.9 (3.6)	16.4 (4)	11.3 (3.2)	11.3 (3.2)
looic	2269.1 (161.5)	2274.6 (162.1)	364.7 (75.7)	364.7 (75.7)

The SCEs were measured as the number of events per 10,000 miles driven. Incidence rate ratio and associated 95% credible intervals were reported.

4.5. Diagnostics statistics and model selection

All the models and truck drivers have Pareto k diagnostic statistics of less than 0.7 (not shown in the tables), which suggests no signal for model misspecification (Vehtari et al., 2015, 2017). The estimated effective number of parameters (`p_loo` in Tables 5, 6, 7), which were similar to the total number of parameters in the models. These two results suggest negative binomial models were reasonably specified models given the large number of observations in this study (Vehtari et al., 2017, 2015). The LOOIC in the tables can be used to compare different models, with lower values indicating better models. In Table 5, the "Four SCEs" model has the lowest LOOIC (79,540.5) among the six models, although the standard errors in the bracket suggested that the "Four SCEs" model was not significantly better than the "Pooled" model.

Figure 4 and 5 present the posterior predictive checks for models in Table 5 and 6. The two figures checked the model prediction accuracy by comparing the observed data to 100 simulated data sets generated from the parameter posterior distributions (Gelman et al., 2013, Chapter 6). For each simulated data set, the proportion of zero crashes was calculated and compared to the observed proportion in the observed data. The black solid

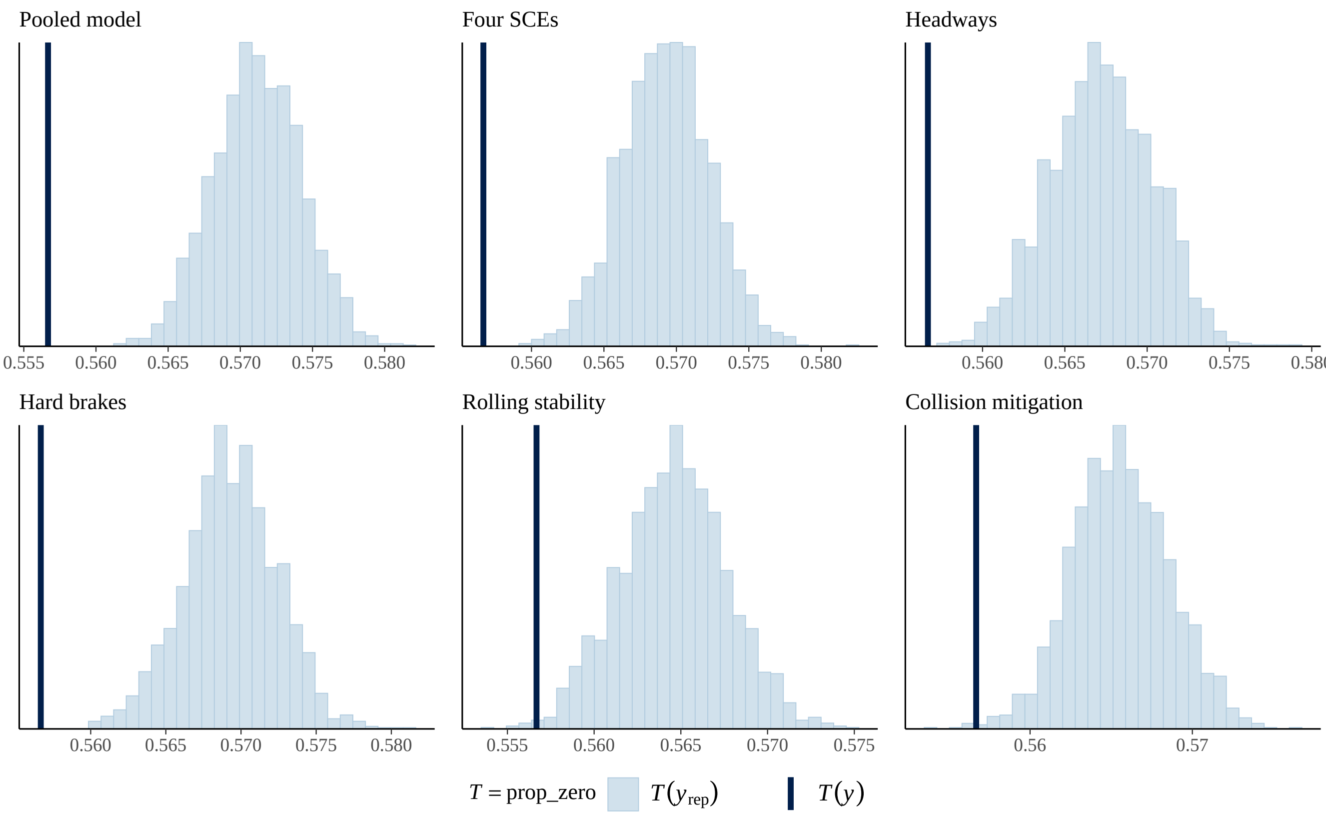


Figure 4: Graphical posterior predictive checks with zero count test statistic for Bayesian negative binomial models for all drivers. The x-axis is the proportion of zero crashes and y-axis is probability density. The solid black line is the observed proportion, while the light blue histogram is from 100 simulated predictions

vertical lines are the observed proportion of zero crashes in observed data, while the light blue histograms show the distribution of proportion of zero crashes in the 100 simulated data sets generated from the parameter posterior distributions. The simulated distributions of zero proportions were on the right of the observed zero proportion in Figure 4, indicating the negative binomial models tend to over-predict non-crashes and under-predict the crashes, but the magnitude of the prediction bias is small. By contrast, in Figure 5, the observed proportion of zero crashes located towards the center of the simulated distributions when stratified by different business units and driver types, which suggested that stratified models in Table 6 have better prediction accuracy and model fit than non-stratified models in Table 5.

5. Discussion

In line with previous studies on the association between crashes and SCE in NDS (Guo et al., 2010, Gordon et al., 2011, Simons-Morton et al., 2012, Wu and Jovanis, 2012, Guo and Fang, 2013, Pande et al., 2017, Gitelman et al., 2018), this study provides statistically significant and robust evidence that there is positive association between SCEs and crashes among commercial truck drivers in the United States. In addition, we have quantified the strength of this association.

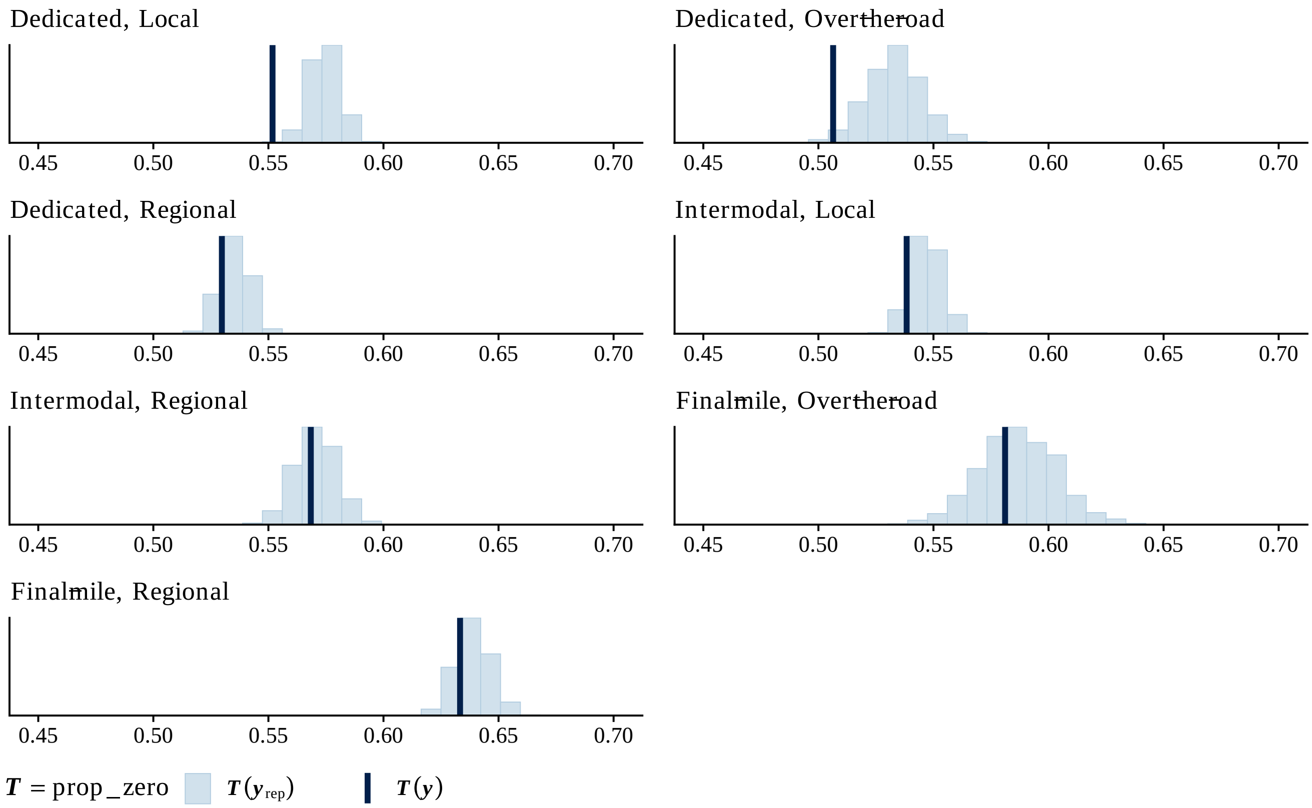


Figure 5: Graphical posterior predictive checks with zero count test statistic for Bayesian negative binomial models, stratified by business unit and driver types. The x-axis is the proportion of zero crashes and y-axis is probability density. The solid black line is the observed proportion, while the light blue histogram is from 100 simulated predictions.

The current study contributes to the existing literature in three aspects. First, this paper overcome the small sample size issues in previous crashes and crash surrogates papers, which typically include no more than 100 drivers or vehicles (Guo et al., 2010, Gitelman et al., 2018). Our study included 31,828 commercial truck drivers, and the large sample size allows us to investigate the association between four different types of SCEs and crashes, as well as stratified analyses across business units and driver types. Second, commercial trucks are the major causes of catastrophic accidents and transportation risk, but evidence on the association between crashes and crash surrogates among truck drivers is scarce. Our study gives insights to this less studied field using a nationwide large-scale sample. Third, this paper explored the association between SCEs and human injuries, which is investigated in previous papers yet important question. We found that as the number of SCEs per 10,000 miles increase by one unit, the number of injuries per mile will increase by 8.7% (95% CI: 4.8%-13.6%).

Among the four types of SCEs studied in this paper, the most frequent one is hard brakes (n = 21,8419), followed by headways (n = 170,421), collision mitigation (n = 55,243), and rolling stability (n = 6,675). The number of hard brakes is around 30 times higher than that of rolling stability. The number of hard brakes and headways are greater than collision mitigation and rolling stability. In statistical analyses, we did not find that hard brakes and headways were over-defined. One unit increase in the number of hard brakes or headways per 10,000 miles were consistently associated with increase in crashes, and this association still holds when stratified by business units and driver types. The incidence rate ratios of hard brakes and headways were smaller than those of collision mitigation and rolling stability, which could be explained by the relatively high frequencies.

Although our study shows a positive association between SCEs and crashes, as well as injuries, we cannot conclude that the second assumption in Heinrich's Triangle is true, i.e., the frequency of severe events can diminish as the frequency of less severe events decreases (Guo et al., 2010). This study used a cross-sectional observatory design and no experiments were involved. Therefore, we should not make any causal statements on this association. Exploring the causal relationship between SCEs and crashes requires well-designed experiments or state-of-start causal inference models for observatory data, which is beyond the scope of this study and could be a direction that future studies can focus on.

Our results are based on truck driving data from a single trucking company, but we argue that the generalizability of the results to other trucking companies is high. First, Company A is a leading trucking and transportation company in the United States, and their business divisions, routing, and scheduling are not particularly special compared to other trucking companies. Second, the commercial trucking industry has a high turn-over rate (Johnson et al., 2010), which means the truck drivers actively switch between companies. Therefore, different trucking companies share a fair amount of drivers and are generally similar regarding drivers. Third, a substantial number of drivers in each of the three business units and three driver types have been included in our study. The number of truck drivers involved is at least 30 times as high as those in other relevant studies, and the traveling pings cover the entire country and most of the major roads. Nonetheless, we still suggest the researchers always investigate the relationship between SCEs and crashes or injuries prior to applying SCEs as surrogates of crashes in risk prediction.

6. Conclusions

6.1. Summary of the Main Contributions

Based on routinely collected NDS and kinematic data from 31,828 truck drivers in a large commercial trucking company, this study investigated the association between SCEs (headways, hard brakes, collision mitigation, and rolling stability) and crashes, injuries, and fatalities using Bayesian negative binomial regression models. We found that one unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% CI: 8-8.8%) increase in SCEs per mile and 8.7% (95% CI: 4.8%-13.6%) increase in the number of injuries per mile. The increase was different in different types of SCEs: 3.3% (95% CI: 2.6-4%) for headways, 8.1% (95% CI: 7.5-8.7%) for hard brakes, 50.4% (95% CI: 41.4-60%) for rolling stability, and 22.2% (95% CI: 19.8-24.5%) for collision mitigation. The results are consistent when stratified by different business units and driver types. This work provides statistically strong and robust evidence that SCEs are positively associated with crashes and injuries among commercial truck drivers.

6.2. Practical Relevance of our Work to the Trucking and General Commuting Populations

Our findings have given evidence to support the long held belief that critical events are associated with crashes. Much of the literature on truck safety has presumed that such an association exists, but often little evidence is provided. Our study is by far the largest one to quantify this association. This result will allow researchers to study the effect of variables on safety by using critical events as a proxy for crashes. Crashes, fortuntely, are somewhat rare, and the ability to use a much more common outcome will give researchers more power when investigating variables that affect safety.

6.3. Limitations and Suggestions for Future Research

This study has several limitations. First, since the exact time of the crashes were not recorded, we are not able to find out which and how many SCEs directly cause crashes, which is a crucial question in Heinrich's Triangle. Second, although we have a relatively large number of drivers and high mileage driven, the number of injuries or fatalities is not sufficiently large for stratified analyses. The statistical evidence on the number of injuries is strong only when we combine four types of SCEs, while the evidence on fatalities is either of the two models is not strong given the small number of fatalities. Third, we do not have data on traffic or weather variables, which are important predictors of crashes. Fourth, since the crash data in this study were reported by the drivers, it is likely that the drivers were underreporting crashes, especially for non-injury or non-fatality injuries.

7. Supplementary Materials

In an effort to promote open and reproducible research, we have published our sample ping data, drivel-level aggregated data[1], an Rmarkdown file including R code and statistical results[2] at a website hosted by GitHub Pages[3]. The original ping data include sensitive latitude and longitude information and cannot be made open. The website contains the following information:

A) Ping data and trip aggregation

- Ping data demonstration
- Aggregating ping data into trips
- Ping and trip data visualization

B) Statistical modeling

- Aggregated driver-level data
- Bayesian Negative binomial regression using `rstanarm`
- Model comparison and diagnostics using `loo`

[1]The data sets are provided in Comma-Separated Values (CSV) files and can be accessed at https://github.com/caimiao0714/Github-SCE-crash/tree/master/data

[2]The Rmarkdown file can be accessed at https://github.com/caimiao0714/Github-SCE-crash/blob/master/index.Rmd

[3]The website can be accessed at https://caimiao0714.github.io/Github-SCE-crash/.

Acknowledgment

This work was supported in part by: the National Science Foundation (CMMI-1635927 and CMMI-1634992); the Ohio Supercomputer Center (PMIU0138 and PMIU0162); the American Society of Safety Professionals (ASSP) Foundation; the University of Cincinnati Education and Research Center Pilot Research Project Training Program; the Transportation Informatics Tier I University Transportation Center (TransInfo); a Google Cloud Platform research grant for data management; and a Dark Sky grant for extended API access of weather data (i.e., they increased the number of allowable queries per day).

References

M. Blanco, R. J. Hanowski, R. L. Olson, J. F. Morgan, S. A. Soccolich, S.-C. Wu, et al. The impact of driving, non-driving work, and rest breaks on driving performance in commercial vehicle operations. Technical report, (DOT-HS-810594). Washington, DC: United States. Federal Motor Carrier Safety Administration, 2011.

C. Chen and Y. Xie. The impacts of multiple rest-break periods on commercial truck driver's crash risk. *Journal of Safety Research*, 48:87–93, 2014.

G. X. Chen, Y. Fang, F. Guo, and R. J. Hanowski. The influence of daily sleep patterns of commercial truck drivers on driving performance. *Accident Analysis & Prevention*, 91:55–63, 2016.

P. Cooper. Experience with traffic conflicts in canada with emphasis on "post encroachment time" techniques. In *International calibration study of traffic conflict techniques*, pages 75–96. Springer, 1984.

T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, et al. The 100-car naturalistic driving study. phase 2: Results of the 100-car field experiment. Technical report, United States. Department of Transportation. National Highway Traffic Safety, 2006.

T. A. Dingus, R. J. Hanowski, and S. G. Klauer. Estimating crash risk. *Ergonomics in Design*, 19(4):8–12, 2011.

R. Eenink, Y. Barnard, M. Baumann, X. Augros, and F. Utesch. Udrive: the european naturalistic driving study. In *Proceedings of Transport Research Arena*. IFSTTAR, 2014.

L. Evans and P. Wasielewski. Do accident-involved drivers exhibit riskier everyday driving behavior? *Accident Analysis & Prevention*, 14(1):57–64, 1982.

L. Evans and P. Wasielewski. Risky driving related to driver and vehicle characteristics. *Accident Analysis & Prevention*, 15(2):121–136, 1983.

Federal Motor Carrier Safety Administration. Large Truck and Bus Crash Facts 2017. https://www.fmcsa.dot.gov/safety/data-and-statistics/large-truck-and-bus-crash-facts-2017, 2018. [Online; accessed 20-June-2019].

J. Fox and G. Monette. Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87 (417):178–183, 1992.

A. E. Gelfand. Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, pages 145–161, 1996.

A. E. Gelfand, D. K. Dey, and H. Chang. Model determination using predictive distributions with implementation via sampling-based methods. Technical report, Department of Statistics, Stanford University, 1992.

A. Gelman, D. B. Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7 (4):457–472, 1992.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013.

V. Gitelman, S. Bekhor, E. Doveh, F. Pesahov, R. Carmel, and S. Morik. Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes. *Transportation Research Part C: Emerging Technologies*, 91:156–175, 2018.

B. Goodrich, J. Gabry, I. Ali, and S. Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2018. URL http://mc-stan.org/. R package version 2.17.4.

T. J. Gordon, L. P. Kostyniuk, P. E. Green, M. A. Barnes, D. Blower, A. D. Blankespoor, and S. E. Bogard. Analysis of crash rates and surrogate events: unified approach. *Transportation research record*, 2237(1):1–9, 2011.

F. Guo. Statistical methods for naturalistic driving studies. *Annual Review of Statistics and Its Application*, 6: 309–328, 2019.

F. Guo and Y. Fang. Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, 61:3–9, 2013.

F. Guo, S. G. Klauer, J. M. Hankey, and T. A. Dingus. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record*, 2147(1):66–74, 2010.

J. M. Hankey, M. A. Perez, and J. A. McClafferty. Description of the shrp 2 naturalistic database and the crash, near-crash, and baseline data sets. Technical report, Virginia Tech Transportation Institute, 2016.

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

R. J. Hanowski, R. Olson, J. Bocanegra, and J. S. Hickman. Analysis of risk as a function of driving-hour: assessment of driving-hours 1 through 11. Technical report, US Department. of Transportation, Federal Motor Carrier Safety Administration, 2008.

J. S. Hickman, R. J. Hanowski, and J. Bocanegra. A synthetic approach to compare the large truck crash causation study and naturalistic driving data. *Accident Analysis & Prevention*, 112:11–14, 2018.

R. J. Hijmans. *geosphere: Spherical Trigonometry*, 2019. URL https://CRAN.R-project.org/package=geosphere. R package version 1.5-10.

C. Hydén. The development of a method for traffic safety evaluation: The swedish traffic conflicts technique. *Bulletin Lund Institute of Technology, Department*, (70), 1987.

J. C. Johnson, D. N. Bristow, D. J. McClure, and K. C. Schneider. Long-distance truck drivers discuss their occupation. *Journal of Transportation Law, Logistics & Policy*, 77(1), 2010.

R. R. Knipling. Naturalistic driving events: No harm, no foul, no validity. 2015.

R. R. Knipling. Threats to scientific validity in truck driver hours-of-service studies. 2017.

D. Lord and F. Mannering. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5):291–305, 2010.

A. Mehdizadeh, M. Cai, Q. Hu, A. Yazdi, M. Ali, N. Mohabbati-Kalejahi, A. Vinel, S. E. Rigdon, K. C. Davis, and F. M. Megahed. A review of data analytic applications in road traffic safety. part 1: Descriptive and predictive modeling. *Sensors*, 20(4):1107, 2020.

V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman. An overview of the 100-car naturalistic study and findings. *National Highway Traffic Safety Administration, Paper*, 5:0400, 2005.

A. Pande, S. Chand, N. Saxena, V. Dixit, J. Loy, B. Wolshon, and J. D. Kent. A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accident Analysis & Prevention*, 101:107–116, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

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

S. Roshandel, Z. Zheng, and S. Washington. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention*, 79:198–211, 2015.

J. H. Saleh, E. A. Saltmarsh, F. M. Favaro, and L. Brevault. Accident precursors, near misses, and warning signs: critical review and formal definitions within the framework of discrete event systems. *Reliability Engineering & System Safety*, 114:148–154, 2013.

P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention*, 43(5): 1666–1676, 2011.

G. Shmueli et al. To explain or to predict? *Statistical science*, 25(3):289–310, 2010.

B. G. Simons-Morton, Z. Zhang, J. C. Jackson, and P. S. Albert. Do elevated gravitational-force events while driving predict crashes and near crashes? *American Journal of Epidemiology*, 175(10):1075–1079, 2012.

R. Sinnott. Virtues of the haversine. sky and telescope. 1984.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

C. Staton, J. Vissoci, E. Gong, N. Toomey, R. Wafula, J. Abdelgadir, Y. Zhou, C. Liu, F. Pei, B. Zick, et al. Road traffic injury prevention initiatives: a systematic review and metasummary of effectiveness in low and middle income countries. *PLoS One*, 11(1):e0144971, 2016.

H. S. Stern, D. Blower, M. L. Cohen, C. A. Czeisler, D. F. Dinges, J. B. Greenhouse, F. Guo, R. J. Hanowski, N. P. Hartenbaum, G. P. Krueger, et al. Data and methods for studying commercial motor vehicle driver fatigue, highway safety and long-term driver health. *Accident Analysis & Prevention*, 126:37–42, 2019.

The WHO. The top 10 causes of death. http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, 2018a. [Online; accessed 20-February-2019].

The WHO. *Global status report on road safety 2018*. World Health Organization, 2018b.

A. Theofilatos, G. Yannis, P. Kopelias, and F. Papadimitriou. Impact of real-time traffic characteristics on crash occurrence: Preliminary results of the case of rare events. *Accident Analysis & Prevention*, 2018.

United States Geological Survey. USGS Small-scale Dataset - 1:1,000,000-Scale Major Roads of the United States 201403 Shapefile. https://www.sciencebase.gov/catalog/item/581d052be4b08da350d524ce, 2014. [Online; accessed 20-September-2019].

A. Vehtari, A. Gelman, and J. Gabry. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2015.

A. Vehtari, A. Gelman, and J. Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27(5):1413–1432, 2017.

A. Vehtari, J. Gabry, Y. Yao, and A. Gelman. loo: Efficient leave-one-out cross-validation and waic for bayesian models, 2019. URL https://CRAN.R-project.org/package=loo. R package version 2.1.0.

C. Wang, M. A. Quddus, and S. G. Ison. The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety Science*, 57:264–275, 2013.

S. Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594, 2010.

K.-F. Wu and P. P. Jovanis. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accident Analysis & Prevention*, 45:507–516, 2012.

K.-F. Wu and P. P. Jovanis. Defining and screening crash surrogate events using naturalistic driving data. *Accident Analysis & Prevention*, 61:10–22, 2013.

K.-F. Wu, J. Aguero-Valverde, and P. P. Jovanis. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis & Prevention*, 72:210–218, 2014.

E. Zaloshnja and T. Miller. Unit costs of medium and heavy trucks. Technical report, Report No. FMCSA-RRA-07-034). Washington, DC: Federal Motor Carrier Safety, 2007.

Z. Zheng, P. Lu, and B. Lantz. Commercial truck crash injury severity analysis using gradient boosting data mining model. *Journal of Safety Research*, 65:115–124, 2018.