Transportation Research Part C

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

--Manuscript Draft--

Manuscript Number:	TRC_2019_1394R1
Article Type:	Research Paper
Keywords:	Safety-critical events; Crashes; Naturalistic driving studies; Injuries; Truck; Fatalities
Corresponding Author:	Steven Rigdon Saint Louis University MO United States
First Author:	Miao Cai
Order of Authors:	Miao Cai
	Mohammad Ali Alamdar Yazdi
	Amir Mehdizadeh
	Qiong Hu
	Alexander Vinel
	Karen Davis
	Hong Xian
	Fadel Megahed
	Steven Rigdon
Abstract:	The past decade has witnessed continuous growth of naturalistic driving studies (NDSs). When analyzing NDS data, safety-critical events (SCEs) are commonly used as surrogates for safety, since actual 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 (headway, hard brake, collision mitigation, and rolling stability) and crashes, as well as injuries and fatalities. Compared to existing studies on the subject, that are based on up to about 2 million miles driven, our study involves an estimated 2.3 billion miles driven. Bayesian negative binomial models were applied to examine the association between three outcomes (crashes, injuries, and fatalities) and the four SCEs. 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): 8.0-8.8%) 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 companies provide a promising opportunity for future data analytic research.
Suggested Reviewers:	Feng Guo feng.guo@vt.edu
	Brian Mayer bmayer@vt.edu Expert in naturalistic driving studies
	Xiao Qin qinx@uwm.edu Expert in traffic operations and intelligent transportation systems

	Michael Fitzharris Michael.Fitzharris@monash.edu Expert in the entire crash-continuum.
Response to Reviewers:	

May 31, 2020
Dr. Yafeng Yin
Editor-in-Chief
Transportation Research Part C: Emerging Technologies
Department of Civil & Environmental Engineering
University of Michigan
2350 Hayward Street
Ann Arbor MI 48109-2125, USA

Dear Dr. Yin,

We are pleased to resubmit our paper "The association between crashes and safety-critical events: synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers" to Transportation Research Part C: Emerging Technologies for review and consideration.

We appreciate the careful reviews provided by the reviewers on the previous submission. We are including our responses to their comments.

The corresponding author of this manuscript will be Steve Rigdon, Ph.D. at steve.rigdon@slu.edu. Thank you in advance for processing our manuscript in a timely fashion.

On behalf of the research team,

Steven E. Rigdon, Ph.D.
Department of Epidemiology and Biostatistics
College for Public Health and Social Justice
Saint Louis University
3545 Lafayette Avenue
St. Louis, MO 63108, USA

Response to Reviewers' Comments for TRC_2019_1394

We would like to thank the three reviewers for their constructive comments, which significantly improved our manuscript. In this document, we detail how we have addressed all the comments made by the reviewers. For the reviewers' convenience, we start our response to each reviewer in a new page. We reproduce a reviewer's comment in a grey box, then provide our response in **bold**. Furthermore, we highlight the changes in the manuscript in **blue**, with the exception of minor edits on typos and language issues.

Response to Reviewer 1's Comments

> The paper reads well and this study explores the association between truck crashes and safety-critical events using crash reports and naturalistic driving data. The topic is interesting. However, the reviewer has some concerns.

Thank you for your careful review of our paper. We appreciate your comments and suggestions. Below, we provide our point-by-point response to your comments.

> 1. Justification of the research gap is weak, especially the introduction. Besides, what are the safety-critical events (SCEs)? Could you provide a definition?

To address your comment, we now integrate the literature review (previously Section 2) with the introduction in a new subsection titled "1.1. Research gaps" To clarify the research gap, we have specifically: (a) added three columns into Table 1 highlighting the data used in the literature, study size, and crash surrogates (and observed effect direction), and (b) included a paragraph explicitly highlighting four main gaps in the literature based on Table 1.

As for the safety critical events (SCEs), we now:

- **Clarify that SCEs are routinely captured in naturalistic driving studies (NDSs):**

  In view of these limitations, naturalistic driving studies often use surrogate events in place of actual crashes. Such surrogates are usually referred to as safety-critical events (SCEs) and are selected to intuitively represent (more numerous) "near misses", i.e., 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).

- **State the sensor-based monitoring system used for collecting the SCEs:**

  At the time of data collection, the company's entire fleet utilized the Bendix® Wingman® Advanced™ monitoring system. Both the data management and system maintenance operations were performed by the company. Note that this system is also widely used by other U.S. based trucking operators.

- **Define the SCEs in Section 1.2:**

  The sensor-based monitoring system used by the company captures four different kinematic events:

  - Headway, which signals an instance of tailgating for at least 118 seconds at an unsafe gap time (a measure of distance between leading and trailing vehicles) of 2.8 seconds or less (Grove et al. 2015).

– Hard brakes, which are defined as instances of a deceleration rate of 9.5 miles per hour per second or more.
– Activation of the rolling stability system, which intervenes by applying brake pressure (in addition to potentially applying trailer pressure) assisting the driver in aligning the vehicle when the system's critical thresholds are approached (Bendix® 2007).
– Activation of the forward collision mitigation system.

2. The introduction is not good and some sentences make them confused. The authors would be suggested to improve it largely.

We have completely rewritten the Introduction. It is currently divided into three main parts: (a) an overview of the importance of trucking safety, how sensor-based technologies can overcome some of the limitations in traditional reporting and the motivation for using SCEs as surrogates for crashes; (b) a subsection detailing the research gaps; and (c) a subsection detailing the study focus, goals and contributions. Based on these substantial changes, we believe that this version is more clear and convincing.

3. Authors also kept silent regarding data quality & integrity.

What is the percentage of drivers excluded?

How about the accuracy of the GPS data? As we know, sometimes, the coordination of the GPS may be far away from the actual location, even the other side of the road.

How to choose the thresholds? Any references?

The quality of crash data in this manuscript should be reported.

There are several aspects that we would like to highlight in how we addressed these comments:

- **Data section: We now include a "Data description" section where we provide some background on the company, summarize the drivers' characteristics, detail the data acquired from the sensor-based monitoring system, and explain our data aggregation process.**

- **Overall data quality considerations: By explicitly stating the name of the sensor-based monitoring system (Bendix® Wingman® Advanced™ monitoring system), we provide researchers with a detailed ability to replicate/test our data collection procedure.**

- **Percentage of excluded drivers: We have added the following description of total number of drivers and the exclusion criteria that we used.**

  The original dataset provided by the company included 34,348 drivers. We have excluded 2,520 drivers (i.e., 7.4% of the original dataset) from our analysis if any of the following criteria is met: (a) driver inactivity, where we required the driver to have at least 100 GPS pings in the data to be included; (b) the unique identification code for the driver is not found in the provided demographics table; and/or

(c) the number of SCEs reported were identified as obvious outliers (we only removed drivers who had an unrealistically large number of SCEs). Hereafter, all reported data will correspond to only those generated by the remaining 31,828 drivers, whose characteristics are summarized in Table 2.

- **Accuracy of GPS data: The GPS information was provided by our industrial partner. The GPS data provided by the Company included a GPS quality indicator, which showed that 98.7% of the data to be of "good quality" for the purposes of vehicle tracking and routing. We do acknowledge that general GPS devices used in motor vehicles are typically accurate within the 3.0 − 4.9 meter range (depending on the device used). However, for our purposes even such general GPS devices are adequate since: (a) GPS data is solely used in estimating the distance traveled in a trip; (b) trips are made up of a large number of pings (79.8 on average); and (c) distance is estimated based on sequence of pings from the beginning to the end of the trip. It is important to note we do not consider typical variables that can be affected significantly with the GPS accuracy, such as (number of lanes, traffic flow, etc.), given that the primary focus of this paper is examining the association between the SCEs and crashes.**

- **Thresholds: We now provide the thresholds for both the headway and hard brake SCEs. We do not provide thresholds for the other two SCEs given that they constitute proprietary information from Bendix®; however, by providing the name of the sensor-based system (and a reference to its user manual) we present all the information needed to capture our data capturing mechanism.**

- **Quality of crash data: The crash data were captured by a leading trucking company. Per the collected data, the company seems to well exceed the reportortable crash guidelines set by the federal government which defines "a reportable crash is one in which a vehicle was towed from the scene, or an injury or fatality occurred." Hence, we consider the crash data to be of high quality. That being said, we also note that a possible limitation of our work is that crashes may be under-reported in Section 5.3.**

4. In 4.1, why not report the driving experience (e.g., driving years)? Why not report the statistical summary of the variables?

We now include tables of the summary statistics for the main predictor variables and the covariates. In Table 2, we provide a summary of driver characteristics including their average age ± SD, number of drivers per gender, business unit and driver types (with their % in parentheses). Furthermore, we provide a summary of the computed response, predictor variables and covariate in Table 4. While the company did provide some data pertaining to "driving years", we elected not to include this in the table of driver demographics since the missing rate of driving years, in the demographics table provided by the company, is more than 30%. From a closer investigation of the column, it seemed that the company started to collect this information recently; drivers who have been employed by the company for over three years typically did not have any information for that variable. Hence, we do not include that variable in our manuscript.

5. Were all variables included in the models?

   So the authors did not consider the multicollinearity?

   And the tables made the reviewer confused.

We used different combinations of predictors and covariates in the different models. To make this point clearer, we have: (a) added a new table (Table 5), where we show how the data are sampled and modeled to examine the three research questions; (b) used different mathematical symbols in equation 2 to distinguish between the main predictors, i.e. the SCEs, and the covariates – we explicitly stated that the number of predictors and covariates will vary; and (c) organized the tables such that they follow the model description process in Table 5.

Yes, we have checked for both the pairwise correlation and multicolinearity in the data. In this revision, we now include the following:

- **In Section 3.1**:

  The final step in preparing the data for modeling was to examine whether the number of predictors and/or covariates can be reduced based on a correlation analysis. Note that the variables included both continuous (primary predictors, age and mean ping speed) and polytomous (gender, business unit, and driver type) variables. To account for the mixed variable types, we adopted the approach of Revelle et al. (2010, 2016) to compute the Pearson, polychoric and polyserial correlation coefficients for the pairwise evaluation of continuous, polytomous and mixed variables, respectively.

  The results of this analysis were presented in Figure 3, which showed that the pairwise correlation among the predictors and covariates is generally small.

- **In Section 3.3.3:**

  Since four different SCEs and/or multiple covariates were included in the models at the same time, it is important to check for the presence of multicollinearity. Recall that in Section 3.1, we examined the correlations among pairs of predictors and found that the pairwise correlations are small. Here, we attempt to investigate whether a linear dependence exists among three or more of our variables through computing the variance inflation factors. If the regressors are uncorrelated, the variance inflation factor obtains its minimum value of 1. In statistical practice, a inflation factor less than 4 requires no additional investigation of linear dependence among the regressors and values greater than 10 indicate serious multicolinearity requiring model corrections.

- **The results from the variance inflation factor computations are now included in Table 9 in Section 4.4.**

  The results capture all three investigated outcomes (crashes, fatalities and injuries) as well as the stratification of the drivers' data set by business unit and driver type. From the table, all variance inflation factors are less than or equal to 1.3. Hence, we can conclude that the variation of each of the regression coefficients are not inflated and that there is no evidence of any multicollinearity issues (Vatcheva et al. 2016).

As for the comment about the confusion from the tables, we believe that the addition of Table 5 and the organization of the results (and Tables 6–8) according to our research questions should clear up any confusion.

6. "In the two models using the number of fatalities as the outcome variable (column 4 and 5), all 95% CIs of IRRs included one and the CIs were very wide", it means that the variables are not statistically significant?

Yes, your observation is correct. To clarify the interpretation of credible intervals in Bayesian inference, we added the following in Section 3.3.2.

In the Bayesian setting, parameters are considered as random variables that have probabilistic distributions instead of unknown fixed values, so no p-values can be reported here. The posterior mean and 95% credible intervals (CIs) of the incidence rate ratios ($\exp(\beta)$) are reported instead. The interpretation of the incidence rate ratios in this Bayesian negative binomial model is as follows: as the number of SCEs per 10,000 miles increases by one unit, the number of crashes per mile is multiplied by $\exp(\beta)$. A 95% credible interval is the interval such that is the posterior probability of the parameter of interest falling within that range given the data is 95% (McElreath 2020) If the 95% CI of the incidence rate ratio includes one, then one is a plausible value for the true incidence rate ratio in this case, and the parameter of the variable will be deemed statistically insignificant. If the 95% credible interval excludes one, then the parameter will be considered as statistically significant.

7. The analysis on the model results, such as the association between four different types of SCEs and crashes, the relationship between the SCEs (e.g., Headways) of the variables, etc., are simple and weak. The authors would be suggested to add the deeper analysis. Otherwise, the contributions of this manuscript would be limited.

In the revision, we have clarified that we ran 17 different Bayesian Negative Binomial models (capturing three different outcome variables, different predictor combinations, and subgroup analyses stratified by different business units and driver types) to address our three research questions of interest. By reorganizing the results section and regression tables according to our three research questions, we allow the reader to map our models to our research questions which address several gaps in the literature as highlighted in Table 1 and Section 1.1. Based on the obtained results, we have shown that our work makes the following contributions to the research and practitioner communities.

Main contributions of the work We've added the following paragraphs:

This work provides statistically strong and robust evidence that SCEs are positively associated with crashes and injuries among commercial truck drivers. Furthermore, this study demonstrates that the "severity" of the SCE is associated with the crash rate, where the two automated maneuvers (involving the initiation of the forward-collision mitigation and rolling stability systems) were shown to have a statistically significant larger effects on crash rates when compared to hard brakes and head way (which can be seen as less severe maneuvers/alerts).

The current study contributes to the existing literature in three respects. First, this paper overcomes the small sample size issues in previous crashes and crash surrogates papers, which typically includes 300 or fewer drivers or vehicles and fewer than 100 crashes (Guo et al. 2010, Gitelman et al. 2018). Our study involves 1,000 times as many driving hours and miles and includes more than $30,000$ commercial truck drivers and $30,000$ crashes; this 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, the evidence of the association between crashes and crash surrogates among truck drivers has been scarce. Our study gives insights to this less studied field using a nationwide large-scale sample. Third, this paper explores the association between SCEs and human injuries and fatalities, which has not been investigated in previous papers but represents important research questions that require detailed study given that they constitute an important component of truck routing models used in practice (Hu et al. 2020).

Practical relevance to trucking operators: We emphasize the practicality of our data collection mechanism, where we use routinely collected data, and discuss why this data is typically untapped by trucking operators. Then, we provide the following three recommendations for how the knowledge of the association between the SCEs and crashes:

First, recent statistics indicate that more than 90% of traffic crashes are influenced by driver behavior (Federal Highway Administration 2019). While our naturalistic driving data does not include video images, it explicitly captures important behavioral factors such as driving speed, aggressive driving through headway alerts, and potential distraction/drowsiness with an increased rate of the three other SCEs (at least when compared to drivers on similar routes). Thus, trucking operators can use our estimated model coefficients, e.g., 50.4% for the initiation of the rolling stability system, in driver training/education. Second, operators can provide incentives to their drivers to reduce their recorded number of SCEs through behavioral based safety programs (Jun et al. 2007). Third, by examining the operator's historical record of SCE data, operators can develop scheduling and routing policies that attempt to minimize the number of recorded SCEs (Mehdizadeh et al. 2020, Hu et al. 2020).

8. The discussion and conclusions would be also suggested to be improved substantially.

We now combine our discussion and conclusions into one section (Section 5), where we discuss: (a) the main contributions of the work, (b) relevance to trucking operations, and (c) pinpointing the limitations in the study and opportunities for future work. In our estimation, the reorganization of those two sections and the added discussion have significantly improved the manuscript. The reviewer is referred to our response to Comment 7 and the revised manuscript for additional details.

9. Many sentences are weak / improper and make readers confused.

   For instance, page 11: 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. How do you know 8.4%?

In the revision, we have made significant changes in the document that should make our paper easier to follow. Pertaining to the specific comment for the coefficient, we can clarify this as follows. Since the number of SCEs per mile $\mu_i = \exp(\alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK})$, the rate change of μ when the primary predictor x_1 is increased by one unit can be calculated as:

$$\begin{aligned}\frac{\mu'}{\mu} &= \frac{\exp[\alpha_0 + \beta_1(x_{i1}+1) + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK}]}{\exp(\alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK})} \\ &= \frac{\exp(\alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK}) \times \exp(\beta_1)}{\exp(\alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK})} \\ &= \exp(\beta_1) \rightarrow \text{which is the incidence rate ratio (IRR).}\end{aligned}$$

Therefore, if the IRR is 1.084, the number of SCEs per mile μ is increased by 8.4% when the predictor x_1 is increased by one unit.

10. The website of the data sets provided in the manuscript does not work.

In this revision, we provide two links in our supplementary materials.

- **Website containing code and analysis: https://caimiao0714.github.io/Github-SCE-crash/.**
- **Sample of masked data: We provide a sample CSV file, where we masked the driver ID as well as the coordinates (by rounding them to one digit) at https://github.com/caimiao0714/Github-SCE-crash/blob/master/data/sample_ping.csv. The purpose of providing the sample file is to provide some insight into the ping data frequency, shape, and speed values.**

These two links were tested on different browsers with no observed issues.

11. Lastly, the review would suggest that the authors re-check their grammar and text as there are many spelling mistakes in the manuscript. For instance, line 2 in page 5: "The found that ......". This manuscript should be revised properly.

The revised paper was checked for grammar and spelling mistakes.

Response to Reviewer 2's Comments

Overall, this is a high-quality paper and is a worthy addition to literatures with some improvement. The study using emerging telematics data and the combination The paper used a large dataset to evaluate the relationship between SCE and crash risk. The topic is certainly worth investigation given the importance of the validity of crash surrogates. Some specific comments are below:

We want to thank you for your careful reading of the paper and the comments and suggestions you have given.

1. Please refrain from using acronyms and abbreviations. There are too many of them and make the paper difficult to follow.

Thank you for your suggestion. We have reduced the frequency of acronyms in our manuscript. Two of the most commonly used acronyms are used throughout the paper: naturalistic driving study (NDS) and safety-critical event (SCE). Other acronyms that have long names if typed out, such as "expected log posterior density leave one out" are used only briefly in Section 3.3; it would be very cumbersome to write (or read) this if we spelled it out each time.

2. Please provide more details on the information about the "ping". For example,
   - Is a ping a single data point or several points?
   - The paper mentioned "ping data every couple of seconds to around 5 minutes." For a particular truck/driver, is the interval fixed or varying?
   - If a vehicle was not on, will a ping still be sent? Figure 1 seems implying so.
   - One important issue is that if a SCE occurred not during at ping period, will it still show in the dataset? Some trigger based system will catch all such events, for example, when the acceleration is above 0.5G, a record will be automatically generated. It is not clear from the description whether this is the case for the data used.

Thank you for your suggestion. Below are our answers to your questions:

- **One ping is one data point giving information about the truck at that time. Variables include latitude, longitude, date, time, real-time speed, and many others.**
- **For a particular truck/driver, the intervals between pings vary.**
- **Yes, the pings are still recorded when a vehicle was not on.**
- **The SCEs were not collected in the vehicle's ping record. The SCEs were collected in a completely different and independent system. These SCEs were collected whenever the kinematic thresholds were triggered by the driver. Therefore, SCE will still be recorded even when there was no ping at that time. The sensor system and kinematic thresholds are introduced in the revised Introduction.**

Our original manuscript did not make it clear that the data were not specifically collected for a safety study. Rather, the data were collected as part of the company's monitoring and surveillance plan. Only afterward was thought given to using the data to assess road safety.

Since we were not involved in the design stage, we had to accept the decisions made regarding the data collection plan. We've added the following paragraph to make this clear.

We must emphasize that these data were collected as part of the company's ongoing monitoring and surveillance plan. The data were not collected specifically as part of a planned NDS. That said, the data are still measurements on factors that could affect safety and are therefore still valid for the purpose of answering the questions we pose in Section 1.2.

> 3. I have major concerns regarding the hard break, which is defined based on the speed decrease: "the speed decrease within a unit time is larger than a preset threshold value." How long is a unit time? I can only specular it is based on the time interval, couple of seconds to around 5 minutes. In reality a hard break typically only takes less than a second. Speed change over more than a few seconds most likely only represents whether the truck got to a stop instead of a hard brake. Can authors clarify how the Hard Brake and headway SCEs were defined. For example, acceleration threshold?

We describe the sensor system used by the company and some of the thresholds in the revised introduction. We've added the following paragraph:

The sensor-based monitoring system used by the company captures four different kinematic events:

- Headway, which signals an instance of tailgating for at least 118 seconds at an unsafe gap time (a measure of distance between leading and trailing vehicles) of 2.8 seconds or less (Grove et al. 2015).
- Hard brakes, which are defined as instances of a deceleration rate of 9.5 miles per hour per second or more.
- Activation of the rolling stability system, which intervenes by applying brake pressure (in addition to potentially applying trailer pressure) assisting the driver in aligning the vehicle when the system's critical thresholds are approached (Bendix® 2007).
- Activation of the forward collision mitigation system.

> 4. For the regression model, I suspect the four SCEs could be correlated. Can you check whether multicolineality is an issue? The results on page 15" "Four SCEs" model was not significantly better than the "Pooled" model." Could be a result of this?

This is an important point that we did not consider in the original manuscript. We have added a correlation matrix plot for the four different types of SCEs (Figure 3 on p. 13). We also conducted a variance inflation factor test for potential multi-collinearity. The Pearson correlation coefficients were all less than 0.2 in magnitude for the four SCEs, and the variance inflation factor scores were all less than 1.3 (Table 9). The only correlation that was even moderately strong was between "Driver Type" and "Ping Speed" which was 0.6. The two tests suggest there was very minor positive correlation among the rates of four types of SCEs, but did not cause significant multicollinearity issues.

5. Please provide appropriate summary statistics for the four SCEs, e.g., mean, standard deviation, and correlation among them. Hard Brake should happen at much higher frequency than other types of SCEs, which could dominate the "All SCEs" variable.

We've added the summary table (Table 4 at the bottom of p. 11) for the whole data set, including the SCEs. We've also added correlation matrix (Figure 3) which indicates that there is only minor correlation between different types of SCEs (Pearson correlation coefficient less than 0.2). In our study, the predictor variables are not the number of SCEs, but the rate of SCEs (the number of SCEs per 10,000 miles). This is because the number of events highly depends on miles driven, while the rate of events is less dependent on miles driven. Table 3 shows that the number of hard brakes, headways, collision mitigation, and rolling stability. The SCEs were dominated by hard brakes (48%) and headways (38%), not by hard brakes alone. Table 4 shows the average rates of SCEs per 10,000 miles were 6.86, 5.35, 0.21, 1.74 for hard brakes, headways, rolling stability, and collision mitigation, and there was large variability in these rates (Table 4).

6. Page 2: "Since NDS data are typically collected every 30 seconds to 5 minutes, the amount of NDS data are generally very large, which provides both an opportunity and a challenge for data analytics." This statement is not accurate. Typical NDS studies used continuous data collection method, e.g., SHRP2 NDS and 100-Car NDS collect data from ignition-on to ignition-off at 10HZ for video and acceleration data.

Thank you for catching this. We have completely rewritten the Introduction and the inaccurate statement has been revised.

7. Figure 3 and 4 are difficult to understand. Why only check the percentage of zeros? Careful exam of Figure 4 show that "the observed proportion of zero crashes located almost exactly at the center of the simulated distributions" is not true: the observed is still far from center for "DED LOC" and "DEC OTR". Overall, I don't think Figure 3 and 4 provide indispensable information on the model performance. The benefit seems to be overwhelmed by the difficulty in understanding them. Unless the authors can substantially improve its clarity.

This is a subtle, but important point. Since we are concerned about how likely our model predicts crashes when the driver has SCEs, comparing the the proportion of zeros in simulated and observed data is informative in prediction accuracy. The purpose is to guard against a model that might fit very well the mean number of SCEs but grossly over-predict (or under-predict) the proportions of zero SCEs. Figures 3 and 4 show that the simulated distributions (light blue histograms) are generally to the right of the observed zero proportion (solid black vertical line), suggesting that our Bayesian negative binomial models tend to over-predict non-crashes but under-predict the crashes, and the magnitude of the prediction bias is small. We have revised the paragraph describing the graphical presentation of the posterior predictive checks in subsection 4.3 and refrained from claiming "the observed proportion of zero crashes located almost exactly at the center of the simulated distributions" in Figure 4. Our revised manuscript says:

To investigate the models' predictive accuracy, we adopt the approach of Gelman et al. (2013, Section 6.3) who suggested simulating some function of the data and parameter, and comparing it with the observed value of a particular quantity. For our trucking safety application, we examined the proportion of zero crashes since it corresponds to a crash-free trip, which is of interest to truck drivers and operators alike. The probability of having zero crashes is, of course, an unknown quantity, but its posterior distribution can be estimated by simulating samples using Hamiltonian Monte Carlo. In this section, we limit our analysis to the models whose outcomes were crashes since the accident and fatality models indicated that our observed events were insufficient for statistical inference (based on the size of the credible interval in Table 6).

Figure 4 shows the posterior distributions, which are indicated by the histograms in light blue, for the posterior probability of zero crashes under each of the six models considered in Table 7. The observed proportion of zero crashes is indicated by the vertical line in each part of Figure 4. For all six models, the observed proportion of zero crashes was considerably less than what would be predicted by the model. Note that the magnitude of this prediction bias is small, usually around 0.015. In other words, while both models (with and without business units and driver types) perform reasonably well in predicting the mean numbers of SCEs, the model with business units and driver types does a better job predicting the proportion of zero crashes. This suggests that different business units and driver types should be accounted for in the model.

Based partly on the result from Figure 4, we ran the model with all four SCEs (model 2) separately for each of the seven business units and driver types. The corresponding posterior predictive check for zero crashes is shown in Figure 5. Here, the vertical lines are much closer to the simulated posterior distribution. This suggests that different business units and driver types should be accounted for in the model.

Minor issues:

1. Equation (1), should subscript "i" be in the right of equation for "y"as well?
2. Page 3: "The large sample size can yield statistically significant results and conclusions." Large sample is a major strength of this paper. However, large sample size does not necessarily yield statistically significant results (if there is no relationship, we probably don't want statistically significant results). More accurate statement is needed, for example: The large sample size provides high statistical power to detect potential relationship between SCEs and crashes.
3. Abstract: second line: should it be "used to measure safety" since "outcome" is a little vague.
4. Figure 3: please label the X-axis. What is PPC in the caption of the figure?

Thank you for these suggestions. We have addressed the issues as suggested.

1. **A subscript "i" has been added to y in the right side of Equation (1).**
2. **The sentence has been changed to: "First, the large sample size (66 million driving hours and 2.3 billion driving miles) provides the statistical power to estimate the correlation between SCEs and crashes."**
3. **The corresponding sentence has been changed to "In NDSs, safety-critical events (SCEs) are commonly used to measure safety since crashes are very rare."**
4. **We have added labels for the horizontal axes in Figure 4 (previously, Figure 3). In addition, we now spell out "posterior predictive checks."**

Response to Reviewer 3's Comments

This paper investigated the association of the surrogate safety metrics and crashes using the NDS data collected from instrumented trucks. Below comments can further improve the quality of the manuscript.

Please contact the journal in case of clarifying the data source and data reliability step.

First, we thank you for your careful reading of the paper and your suggestions for improvement. We have clarified the data sources and quality of the collected data in the revised Data section. Due to the proprietary nature of the data, we cannot provide the actual data in the supplementary materials. Instead, we provide a hypothetical data set similar to our ping data in our supplementary materials section.

While authors reviewed some NDS studies, the biggest NDS study in the US and Europe i.e. "SHRP2" and UDRIVE are missing from the review. Below papers can provide more info about the latest and the largest scale NDS studies with more than 3200 drivers. Please update the Table 1, accordingly.

- Complementary Methodologies to Identify Weather Conditions in Naturalistic Driving Study Trips: Lessons Learned from the SHRP2 Naturalistic Driving Study & Roadway Information Database
- Eenink, R., Barnard, Y., Baumann, M., Augros, X., & Utesch, F. (2014). UDRIVE: the European naturalistic driving study. In Proceedings of Transport Research Arena. IFSTTAR.
- The impacts of heavy rain on speed and headway behaviors: an investigation using the SHRP2 naturalistic driving study data (TRC).
- The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters

We have read through the provided three references. These papers did use large NDS data sets, but they did not investigate the association between crashes and SCEs like those papers listed in Table 1. We were careful in the revision to indicate that the table describes studies on the relationships between SCEs and crashes. Here are a brief summary of the recommended papers:

1. Ghasemzadeh et al. (2019) used the SHRP2 NDS database with more than 3,500 drivers and described three methods to merge NDS data sets with weather information: wiper status, National Climate Data Center, and weather-related crashes. They also provided data reduction and processing procedures for the SHRP2 NDS data. The paper did not investigate the relationship between crashes and SCEs.
2. Eenink et al. (2014) and Barnard et al. (2016) introduces the overarching goal, methodology, research questions, data to be collected, and expected outcomes of the European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment (UDRIVE) project. This paper is more like a research proposal than an evaluation of the data, and the association between crashes and SCEs is a proposed research question but unanswered.
3. Ahmed & Ghasemzadeh (2018) is a data analytics paper that quantifies the association between weather conditions and driver speed and headway selection behaviors using the SHRP2 NDS database. They did not investigate the association between driver speed, headway selection, and real crashes.

Please introduce the data ping. What frequency of data a data ping is representing?

Thank you for your suggestion. We have updated our description of the data ping, including the frequency of the times between pings in the revised manuscript:

Our study is based on data captured from April 1, 2015 to March 31, 2016 by the company's sensor-based monitoring system on their entire fleet. Our dataset includes intermittently collected real-time driving *ping* records, which ranges from every couple of seconds to approximately 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 patterns explaining the variations in interval lengths. 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 with five decimal place recordings), GPS quality, speed, and drivers' anonymized unique identification code.

We also pointed out that the data set we used for this study was collected for a different purpose, although we argue that it still provides safety information. Only afterward was thought given to using the data to assess road safety. Because of this, we were unable to have any influence on the design of the study regarding things like choosing the ping intervals. We have added this paragraph in Section 2.2.1:

We must emphasize that these data were collected as part of the company's ongoing monitoring and surveillance plan. The data were not collected specifically as part of a planned NDS. That said, the data are still measurements on factors that could affect safety and are therefore still valid for the purpose of answering the questions we pose in Section 1.2.

In addition, a hypothetical sample of ping data (in comma separated values format) is provided in our supplementary materials. The readers are encouraged to check out the ping data if they feel unclear/confused by our description.

- Please explain the collision mitigation surrogate in Table 1.
- The authors should explain the method that they calculated headway. Were the vehicles instrumented with radar?
- More explanation of the hard brakes and the threshold that was used should be added. What was the threshold for the 231101 hard brakes? Clarify whether this number represents the events or data pings.
- Rolling stability should be defined.
- Description of the headway calculation and the threshold for critical headway SCE should be added.
- What are the present thresholds on page 8?

The definitions of SCEs were determined by our partner company as a part of their own routine monitoring program using the Bendix® Wingman® Advanced™ monitoring system. The definitions of the four SCEs are defined in the revised Introduction section.

- Headway, which signals an instance of tailgating for at least 118 seconds at an unsafe gap time (a measure of distance between leading and trailing vehicles) of 2.8 seconds or less (Grove et al. 2015).

- Hard brakes, which are defined as instances of a deceleration rate of 9.5 miles per hour per second or more.
- Activation of the rolling stability system, which intervenes by applying brake pressure (in addition to potentially applying trailer pressure) assisting the driver in aligning the vehicle when the system's critical thresholds are approached (Bendix® 2007).
- Activation of the forward collision mitigation system.

> Regarding rolling stability, any NDS study that consider adverse weather and driver performance?

This is a good point, but it is not central to the primary question of our study: the relationships between SCEs and crashes. We have not found any research articles that study the the association between rolling stability and crashes controlling for adverse weather and driver performance. The relationship between rolling stability, as well as the other three SCEs, and other factors such as weather, traffic, driver characteristics, etc., is a topic for future research.

> It would be interesting to see challenges with the data, missing values, etc. to be explained in a paragraph as a data preparation stage.

This is an excellent point. With such a large file (1.2 tera-bytes) data management and cleaning was quite a task. We have added a more complete discussion of missing values and data quality control in the first paragraph of Data Description subsection:

The original dataset provided by the company included 34,348 drivers. We have excluded 2,520 drivers (i.e., 7.4% of the original dataset) from our analysis if any of the following criteria is met: (a) driver inactivity, where we required the driver to have at least 100 GPS pings in the data to be included; (b) the unique identification code for the driver is not found in the provided demographics table; and/or (c) the number of SCEs reported were identified as obvious outliers (we only removed drivers who had an unrealistically large number of SCEs). Hereafter, all reported data will correspond to only those generated by the remaining 31,828 drivers, whose characteristics are summarized in Table 2.

> Page 9 the authors mentioned the median distance of the trip and the median number of miles per trip as 2.61 and 77.06. Did the authors only considered trucks in the urban environment?

This data set covers the national truck transportation environment in the United States. Figure 2 shows the geographic spread of the pings. Although most of the ping data were in urban areas, rural areas were also covered. In fact, in rural areas, the trucks were probably on interstate highways where the speed limit is higher. With the same distribution of times between pings, the trucks would seem to ping less often in rural areas because they cover more ground in that amount of time. Note that here by "trip" we refer to continuous driving with no stops longer than 30 minutes, which is independent from trucks origin or destination. The actual dispatch and driver's shift in our data typically consist of a number of "trips" pieced together, separated by, for example, rest stops.

Xk should be xik

We have corrected this subscript issue.

Authors need to explain how did they come up with K values.

We did not propose these Pareto k values. They were proposed by (Vehtari et al. 2015, 2017). We have added citations in the Methods and Results section to clarify this.

It is recommended that page 10 paragraph 1 be summarized in a table and provide stat for each category.

A summary table (Table 3) has been added at the bottom of p. 11.

Page 11 talked about table 2 and table two is presented in page 14. Please keep the tables close to the description, if possible.

We have reformatted the manuscript to keep the tables and figures closer to where they are referenced. We used LaTeXto typeset the document, and sometimes it has a mind of its own about where tables and figures should go. By overriding some of the default settings, we were able to place the references closer to the objects.

References

Ahmed, M. M. & Ghasemzadeh, A. (2018), 'The impacts of heavy rain on speed and headway behaviors: an investigation using the shrp2 naturalistic driving study data', *Transportation Research Part C: Emerging Technologies* **91**, 371–384.

Barnard, Y., Utesch, F., van Nes, N., Eenink, R. & Baumann, M. (2016), 'The study design of udrive: the naturalistic driving study across europe for cars, trucks and scooters', *European Transport Research Review* **8**(2), 14.

Bendix® (2007), 'Bendix® ABS-6 Advanced with ESP® Stability System - frequently asked questions to help you make an intelligent investment in stability', Bendix Commercial Vehicle Systems LLC, a member of the Knorr-Bremse Group. https://www.bendix.com/media/documents/products_1/absstability/truckstractors/StabilityFAQ.pdf. [Published March 2007; accessed April 19, 2020].

Dingus, T. A., Hanowski, R. J. & Klauer, S. G. (2011), 'Estimating crash risk', *Ergonomics in Design* **19**(4), 8–12.

Eenink, R., Barnard, Y., Baumann, M., Augros, X. & Utesch, F. (2014), UDRIVE: the European naturalistic driving study, *in* 'Proceedings of Transport Research Arena', TRA 2014, 14-17 Apr 2014, Paris, France. IFSTTAR.
URL: *http://eprints.whiterose.ac.uk/93078/*

Federal Highway Administration (2019), 'Human factors', U.S. Department of Transportation, https://highways.dot.gov/research/research-programs/safety/human-factors. [Updated December 2, 2019; accessed April 29, 2020].

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

Ghasemzadeh, A., Hammit, B. E., Ahmed, M. M. & Eldeeb, H. (2019), 'Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the shrp2 naturalistic driving study & roadway information database', *Safety Science* **119**, 21–28.

Gitelman, V., Bekhor, S., Doveh, E., Pesahov, F., Carmel, R. & Morik, S. (2018), '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.

Grove, K., Atwood, J., Hill, P., Fitch, G., DiFonzo, A., Marchese, M. & Blanco, M. (2015), 'Commercial motor vehicle driver performance with adaptive cruise control in adverse weather', *Procedia Manufacturing* **3**, 2777–2783.

Guo, F., Klauer, S. G., Hankey, J. M. & Dingus, T. A. (2010), 'Near crashes as crash surrogate for naturalistic driving studies', *Transportation Research Record* **2147**(1), 66–74.

Hu, Q., Cai, M., Mohabbati-Kalejahi, N., Mehdizadeh, A., Yazdi, A., Ali, M., Vinel, A., Rigdon, S. E., Davis, K. C. & Megahed, F. M. (2020), 'A review of data analytic applications in road traffic safety. part 2: Prescriptive modeling', *Sensors* **20**(4), 1096.

Jun, J., Ogle, J. & Guensler, R. (2007), 'Relationships between crash involvement and temporal-spatial driving behavior activity patterns: use of data for vehicles with global positioning systems', *Transportation Research Record* **2019**(1), 246–255.

McElreath, R. (2020), *Statistical rethinking: A Bayesian course with examples in R and Stan*, CRC press.

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

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A. & Elleman, L. G. (2016), *Web and phone based data collection using planned missing designs*, Sage Publications, Inc, pp. 578–595.

Revelle, W., Wilt, J. & Rosenthal, A. (2010), *Individual Differences in Cognition: New Methods for Examining the Personality-Cognition Link*, Springer New York, New York, NY, pp. 27–49.

Saleh, J. H., Saltmarsh, E. A., Favaro, F. M. & Brevault, L. (2013), '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.

Vatcheva, K. P., Lee, M., McCormick, J. B. & Rahbar, M. H. (2016), 'Multicollinearity in regression analyses conducted in epidemiologic studies', *Epidemiology (Sunnyvale, Calif.)* **6**(2).

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

Vehtari, A., Simpson, D., Gelman, A., Yao, Y. & Gabry, J. (2015), 'Pareto smoothed importance sampling'.

Highlights:

- Safety-critical events are associated with the risk of crashes and injuries.
- The positive association held when stratified by business unit and driver type.
- We analyzed over 1.4 billion pings from 31,690 commercial truck drivers.
- Rolling stability has the highest correlation with crashes.
- Our code, data, and analytic methods are available to promote future research.

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], Hong Xian[a], Fadel M. Megahed[e], 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]*Farmer School of Business, Miami University, Oxford, OH, 45056, United States*

Abstract

The past decade has witnessed continuous growth of naturalistic driving studies (NDSs). When analysing NDS data, safety-critical events (SCEs) are commonly used as surrogates for safety, since actual 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 (headway, hard brake, collision mitigation, and rolling stability) and crashes, as well as injuries and fatalities. Compared to existing studies on the subject, that are based on up to about 2 million miles driven, our study involves an estimated 2.3 billion miles driven. Bayesian negative binomial models were applied to examine the association between three outcomes (crashes, injuries, and fatalities) and the four SCEs. 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): 8.0-8.8%) 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 companies provide a promising opportunity for future data analytic research.

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

*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), hong.xian@slu.edu (Hong Xian), fmegahed@miamioh.edu (Fadel M. Megahed), steve.rigdon@slu.edu (Steven E. Rigdon)

1. Introduction

Truck drivers "form the lifeblood of [the U.S.] economy" (The White House, 2020), moving more than 70% of goods, transporting approximately 10.8 billion tons of freight, and generating annual revenues of over $700 billion (John, 2019). It is then imperative to reduce the prevalence and severity of crashes involving commercial trucks. Tsai et al. (2018) pointed out three aspects related to this concern: (a) in 2017, 5,005 lives were claimed in 4,455 fatal crashes involving large trucks or buses, and 170,000 persons were injured in 116,000 injury crashes involving trucks and buses (FMCSA, 2019); (b) larger property damage occurred, which can increase the likelihood of injuries and human loss, as well as increase travel-time for other commuters; and (c) indirect losses in efficiency arose from the slowing and damage of essential goods during transit. Despite the significant advances in trucking safety technologies (e.g., deployment of forward-collision mitigation and lane departure warning systems), large truck crash rates have increased over the past decade. In the U.S., the involvement rate per 100 million large-truck miles traveled increased from 1.32 in 2008 to 1.48 in 2016 for fatal crashes, and from 21 in 2008 to 31 in 2015 (most recent data at the time of writing) for injury crashes (NHTSA's National Center for Statistics and Analysis, 2019).

Traditionally, the literature focusing on analyzing trucking crashes relied on data from police reports. Due to the retrospective nature of these reports, they are subject to four major limitations. First, the post-hoc nature of the reports results in recall and information bias (Stern et al., 2019). The second limitation stems from the fact that crashes constitute rare events when compared to non-crashes, which makes it difficult to make statistically valid conclusions on the risk factors of crashes (Guo et al., 2010; Theofilatos et al., 2019). Third, critical behavioral factors, such as fatigue, distraction and cell-phone use, are under-reported by those involved in crashes in an attempt to avoid penalties, fines and legal actions (Dingus et al., 2011; Stern et al., 2019). Finally, minor crashes, resulting in no injuries and no/minor physical damages, tend to be under-reported (Blower, 2017), which impacts the total reported crashes and hence, the understanding of crash risk.

The emergence of sensor-based monitoring technologies, which generate detailed data related to both the truck and truck driver's performance, can enhance our understanding of road safety especially when integrated with data analytic techniques (Shi and Abdel-Aty, 2015; Katrakazas et al., 2015; Imprialou and Quddus, 2019; He et al., 2019; Wali and Khattak, 2020; Mehdizadeh et al., 2020; Hu et al., 2020). The coupling of these sensors with computational advances, which can harness the resulting big datasets, have allowed for conducting several naturalistic driving studies (NDSs) that can overcome the aforementioned limitations in the retrospective crash analysis literature. For example, the 100-car study (Neale et al., 2005), UDRIVE (Eenink et al., 2014) and SHRP2 (Hankey et al., 2016) NDSs collected acceleration, video and location data for crashes and non-crashes, thus, providing superior ability to estimate the risk or rate of crashes due to the inclusion of non-crashes, and also to investigate the short period prior to crashes or safety events without being subjected to information/recall bias (Guo et al., 2010).

Despite the massive amounts of data captured in NDSs, the recorded number of crashes, in any given study, remains small. For example, the 100-car study captured around two million vehicle miles, 40,000+ hours of driving, and yet included only 69 crashes (Neale et al., 2005). This is an expected result since crashes are rare events; hence their rates are reported per 100 million miles driven. In view of these limitations, naturalistic driving studies often use surrogate events in place of actual crashes. Such surrogates are usually referred to as safety-critical events (SCEs) and are selected to intuitively represent (more numerous) "near misses", i.e., 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). A fundamental question then is whether some specific classes of SCEs (e.g., hard brakes, initiation of collision mitigation systems, rolling stability alerts, etc.) can be used as a surrogate for crashes. Since these less severe events are more frequent than crashes, a strong positive correlation between the two would allow for statistically valid modeling of crash risk even when the number of actual crashes is limited.

1.1. Research gaps

The study of the association between surrogate measures and more rare accident events can be first attributed to Heinrich (1931) who attempted to apply the scientific method to the field of industrial accident prevention. In the context of motor vehicle crash analysis, the earlier works focused on estimating the conversion factor between the two types of events (Evans and Wasielewski, 1982, 1983; Cooper, 1984; Risser, 1985; Hydén, 1987). With the advancements in sensor-based technologies and the transition to NDS, we found several research studies (Dingus et al., 2006; Guo et al., 2010; Gordon et al., 2011; Simons-Morton et al., 2012; Wu and Jovanis, 2012; Guo and Fang, 2013; Wu et al., 2014; Pande et al., 2017; Gitelman et al., 2018) that examined the relationship between kinematic-based SCEs and crashes. Table 1 provides an overview of these studies on the relationship between SCEs and crashes, highlighting the sample size, driver types, driving locations, number of observed crash and surrogate events for the participating vehicles/drivers, statistical approach used, and statistically significant effects. While the largest of these involved about 43,000 hours and 2 million miles driven, ours is based on 66 million hours and 2.3 billion miles driven.

Table 1: Differences between existing literature and our study regarding sample size, crash surrogates, statistical models, and conclusion.

Ref.	**Data set**	**Study size**	**Region**	**Statistical model**	**Crash surrogates (effect direction)**
Dingus et al. (2006)	100-car	Drivers: 241 commuters Hours driven: 43,000 Miles driven: ∼2M Time frame: 1 year Crashes: 69 Surrogates: 761	Northern Virginia & Washington D.C., USA	95% confidence limits modeled using a Poisson distribution	Braking (↑) Steering (↑) Accelerating (↑)
Guo et al. (2010)	100-car	Drivers: 241 commuters Hours driven: 43,000 Miles driven: ∼2M Time frame: 1 year Crashes: 69 Surrogates: 761	Northern Virginia & Washington D.C., USA	Sequential factor analysis, Poisson regression	Near crashes (↑)
Gordon et al. (2011)	-	Drivers: 78 commuters Hours driven: NR Miles driven: ∼0.08M Time frame: 10 months Crashes: NR Surrogates: NR	Michigan, USA	Seemingly unrelated regression, Poisson regression	Lateral deviation (↓) Lane-departure warning (↑) Time to edge crossing (↑)
Simons-Morton et al. (2012)	-	Drivers: 42 newly licensed teenagers Hours driven: NR Miles driven: NR Time frame: 18 months Crashes: 37 Surrogates: NR	Virginia, USA	Logistic regression using generalized estimating equations	Elevated gravitational-force (↑)
Wu and Jovanis (2012)	100-car	Drivers: 241 commuters Hours driven: 43,000 Miles driven: ∼2M Time frame: 1 year Crashes: 13 Surrogates: 38	Northern Virginia & Washington D.C., USA	Logistic regression	Yaw rate (↓) Lateral acceleration (↓)
Guo and Fang (2013)	100-car	Drivers: 102 young and high mileage Hours driven: 43,000 Miles driven: ∼2M Time frame: 1 year Crashes: 60 Surrogates: 7,394	Northern Virginia & Washington D.C., USA	Negative binomial regression	Critical-incident events (↑)
Wu et al. (2014)	100-car	Drivers: 90 commuters Hours driven: NR Miles driven: ∼1.1M Time frame: 1 year Crashes: 14 Surrogates: 182	Northern Virginia & Washington D.C., USA	Poisson regression	Run-off-road events (↑)

Continued on next page

Table 1 – continued from previous page

Ref.	Data set	Study size	Region	Statistical model	Crash surrogates (effect direction)
Pande et al. (2017)	-	Drivers: 33 commuters Hours driven: NR Miles driven: NR Time frame: 10 days Crashes: NA Surrogates: NA	California, USA	Negative binomial regression	High magnitude jerks while decelerating (↑)
Gitelman et al. (2018)	-	Drivers: 64 commuters Hours driven: NR Miles driven: NR Time frame: 1 year Crashes: NA Surrogates: NA	Israel	Negative binomial regression	Braking (↑) Speed alerts (↓)
This paper	-	Drivers: 31,828 commercial truck drivers Hours driven: ∼66M Miles driven: ∼2.3B Time frame: 1 year Crashes: 34,884 Surrogates: 450,758	National data, USA	Bayesian negative binomial regression	Headway (↑) Hard brakes (↑) Collision mitigation (↑) Rolling stability (↑)

Abbreviations: NR indicates that the parameter was not reported (or reported in combination with another parameter and hence cannot be inferred). M and B denote that the reported numbers are in millions and billions, respectively.

Although the studies summarized in Table 1 use different types of SCEs, they generally reached similar conclusions indicating that their selected crash surrogates had positive or zero association with crashes (with the exception of speed alert events in Gitelman et al. (2018)). Based on Table 1, we can identify four main gaps in the literature:

- Although two separate NDS datasets, 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. It is unclear whether the associations identified for a general driving population can hold for commercial drivers.
- The number of drivers reported in those studies are limited, with the largest studies (Dingus et al., 2006; Guo et al., 2010) examining 241 drivers. Thus, the number of reported crashes was less than 100. Note this observation specifically pertains to the literature examining the association between crash and SCEs since the number of reported crashes in the overarching NDS study can be larger.
- With more than 30,000 truck drivers and 2 billion mileage driven, we can further explore the correlation between SCEs and crash severity (injuries and fatalities), which no existing studies were able to analyze due to small sample size.

- The studies investigating the association between surrogates and crashes were confined to small geographic areas, which may limit the generalizability of the conclusions to regions having different weather conditions, regulations, and commuting behavior (Tsai et al., 2015).

1.2. *Study focus, goals and contributions*

Owing to the four identified gaps in the literature, the primary objective of this paper is to investigate the association between crashes and SCEs among commercial truck drivers. This study is performed in collaboration with a leading shipping freight company in the U.S. The collaboration with an industry partner provides the following unique settings:

- At the time of data collection, the company's entire fleet utilized the Bendix® Wingman® Advanced™ monitoring system. Both the data management and system maintenance operations were performed by the company. Note that this system is also widely used by other U.S. based trucking operators.
- The truck drivers included in this study were all employed by the company at the time of data collection. The data were collected routinely as a part of their job.
- Unlike previous studies, the routes chosen by the drivers are subject to company policies, delivery windows and government regulations, i.e., naturally follow realistic commercial driving patterns.

Consequently, we can capitalize on data generated by 30,000+ trucks, which accumulate over 2 billion total miles per year. The implication of the large sample size is two-fold. First, we can overcome the sample size and driving location restrictions that can affect the inference from the previous studies. Second, the large sample size provides an opportunity for not only studying the relationship between the surrogates and crashes, but to also study the relationship between the surrogates and both injuries and fatalities.

This study investigates whether kinematic events captured from a popular, commercial sensor-based system can be used as surrogates for crash events. If such a relationship can be established, then our insights can be incorporated by other trucking companies employing (similar) systems. The sensor-based monitoring system used by the company captures four different kinematic events:

- Headway, which signals an instance of tailgating for at least 118 seconds at an unsafe gap time (a measure of distance between leading and trailing vehicles) of 2.8 seconds or less (Grove et al., 2015).
- Hard brakes, which are defined as instances of a deceleration rate of 9.5 miles per hour per second or more.
- Activation of the rolling stability system, which intervenes by applying brake pressure (in addition to potentially applying trailer pressure) assisting the driver in aligning the vehicle when the system's critical thresholds are approached (Bendix®, 2007).
- Activation of the forward collision mitigation system.

Our goal is to explain the association between these four SCEs and trucking crashes, injuries and fatalities. To achieve this goal, we examine the following research questions:

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

(B) Is the association between crashes and SCEs consistent for all four different SCEs (headway alerts, hard brakes detected, collision mitigation activation, and rolling stability activation)?

(C) Does the association between crashes and SCEs hold for the different business units of our industrial partner (dedicated, intermodal, and final-mile) and driver types (local, regional, and over-the-road)?

Note that the question of what factors affect the likelihood of a crash, or predict crashes is a separate topic, (Shmueli et al., 2010) which we do not address here.

By addressing the research questions above, this paper provides three main contributions to the literature. First, the large sample size provides the statistical power to estimate the correlation between SCEs and crashes. Second, four types of SCEs (headways, hard brakes, collision mitigation, and rolling stability) are examined in this study, and we show that the magnitude of associations 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 usually too small to infer an association between SCEs and severe crashes. In this data set, we are able to examine the association between injuries, fatalities, and SCEs.

2. Data description

The data used in this study were collected by a U.S.-based trucking company, which provides transportation, delivery, and logistics services to customers and consumers in North America. The data captures three of the logistical services provided by the company: (a) dedicated contract carriage, where trucks and drivers are assigned to a singular customer with familiar routes, task and work duties; (b) intermodal freight services, where the freight is transported in intermodal containers between shipping ports, rail terminals and inland shipping docks; and (c) final mile delivery services, where non-conveyable products are delivered to customers. It is important to note that the purposes of data collection, which covers the company's entire fleet, were routine performance monitoring, drivers' assistance and regulatory compliance. All the data were anonymized prior to the research team's access. The study protocol and data usage for research purposes were approved by the Institutional Review Board of Saint Louis University.

2.1. Drivers' characteristics

The company provided a driver's characteristics table that linked age, gender, business unit and driver type to the anonymized driver ID. Driver types include (a) local drivers who transport freight within a 200-mile radius and return home on the same day, (b) regional drivers moving freights in regional routes that may include several surrounding states, and (c) over-the-road drivers who specialize in hauling freight long distances, requiring them to be on the road for days/weeks.

The original dataset provided by the company included 34,348 drivers. We have excluded 2,520 drivers (i.e., 7.4% of the original dataset) from our analysis if any of the following criteria is met: (a) driver inactivity, where we required the driver to have at least 100 GPS pings in the data to be included; (b) the unique identification code for the driver is not found in the provided demographics table; and/or (c) the number of SCEs reported were identified as obvious outliers (we only removed drivers who had an unrealistically large number of SCEs). Hereafter, all reported data will correspond to only those generated by the remaining 31,828 drivers, whose characteristics are summarized in Table 2.

Table 2: A summary of driver characteristics including their average age ± SD, number of drivers per gender, business unit and driver types (with their % in parentheses).

Variable	Statistics
Age:	
Range	20 to 82 years
Mean age ± SD	44.48 ± 11.72
Gender: (%)	
Male	29,248 (91.9%)
Female	1,583 (5.0%)
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%)
Regional	15,707 (49.3%)
Over-the-road	2,740 (8.6%)

2.2. *Data acquired from the sensor-based monitoring system*

2.2.1. *Ping data*

Our study is based on data captured from April 1, 2015 to March 31, 2016 by the company's sensor-based monitoring system on their entire fleet. Our dataset includes intermittently collected real-time driving *ping* records, which ranges from every couple of seconds to approximately 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 patterns explaining the variations in interval lengths. 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 with five decimal place recordings), GPS quality, speed, and drivers' anonymized unique identification code.

We must emphasize that these data were collected as part of the company's ongoing monitoring and surveillance plan. The data were not collected specifically as part of a planned NDS. That said, the data are still measurements on factors that could affect safety and are therefore still valid for the purpose of answering the questions we pose in Section 1.2.

The included 31,828 commercial truck drivers had a total of $1,494,678,173$ pings. Based on the GPS quality indicator, 98.7% of these pings were of "good quality". An overview of the pings' locations is shown in Figure 1, where the active (speed >0 MPH) and inactive (speed $=0$ MPH) pings are depicted in Figures 1a and 1b, respectively. Note that the plot was created by aggregating the GPS locations to two decimal places in order to ensure the anonymity of the data, and only displaying locations with at least fifty pings. Both Figures 1a and 1b utilize a sequential color scheme, where a darker color indicates a 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 closely matches the U.S. population density distribution (i.e., it is more concentrated along the coasts). The active and inactive pings are generally consistent, but active pings are more concentrated in major midwest roads.

2.2.2. *Safety critical events*

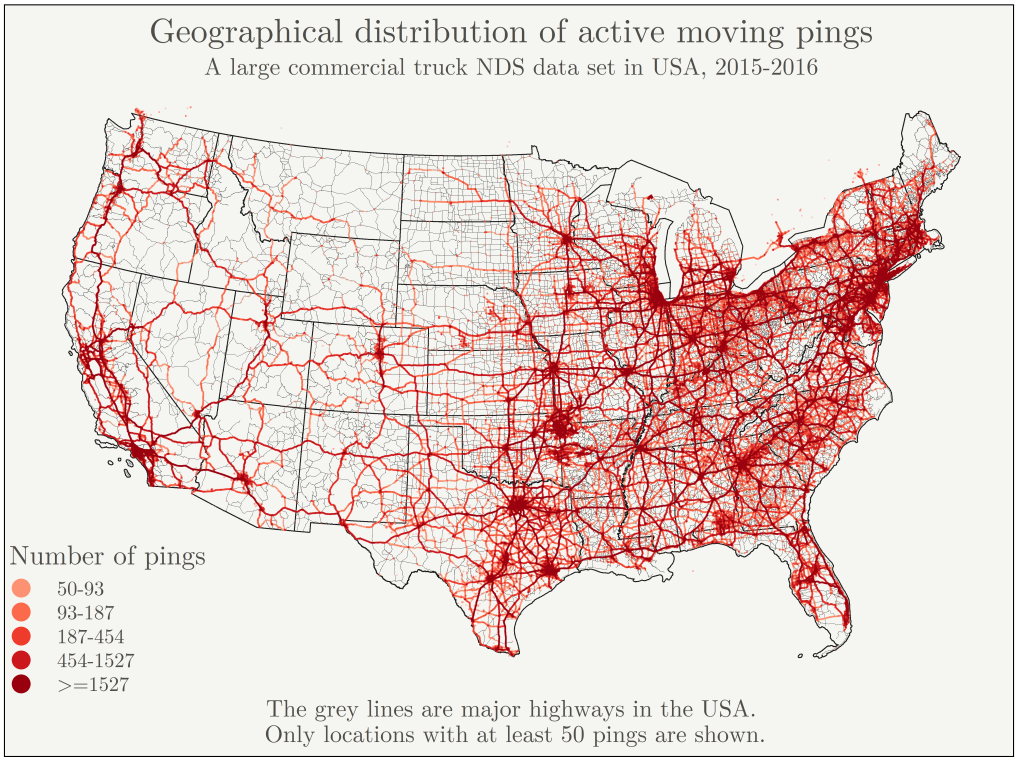
The SCEs were collected independently from the ping data. A SCE is recorded whenever a pre-specified kinematic threshold, defined in Section 1.2, of the Bendix® Wingman® Advanced™ monitoring system was triggered while driving. The collected SCE data include the exact date and time, latitude and longitude (specific to five decimal places), driver, and the type of SCEs. A total of $450,758$ SCEs were collected, which were divided into (a) $170,421$ headway events, (b) $218,419$ hard brakes, (c) $55,243$ initiations of the forward-collision mitigation system, and (d) $6,675$ initiations of the rolling stability system.

2.3. *Crashes*

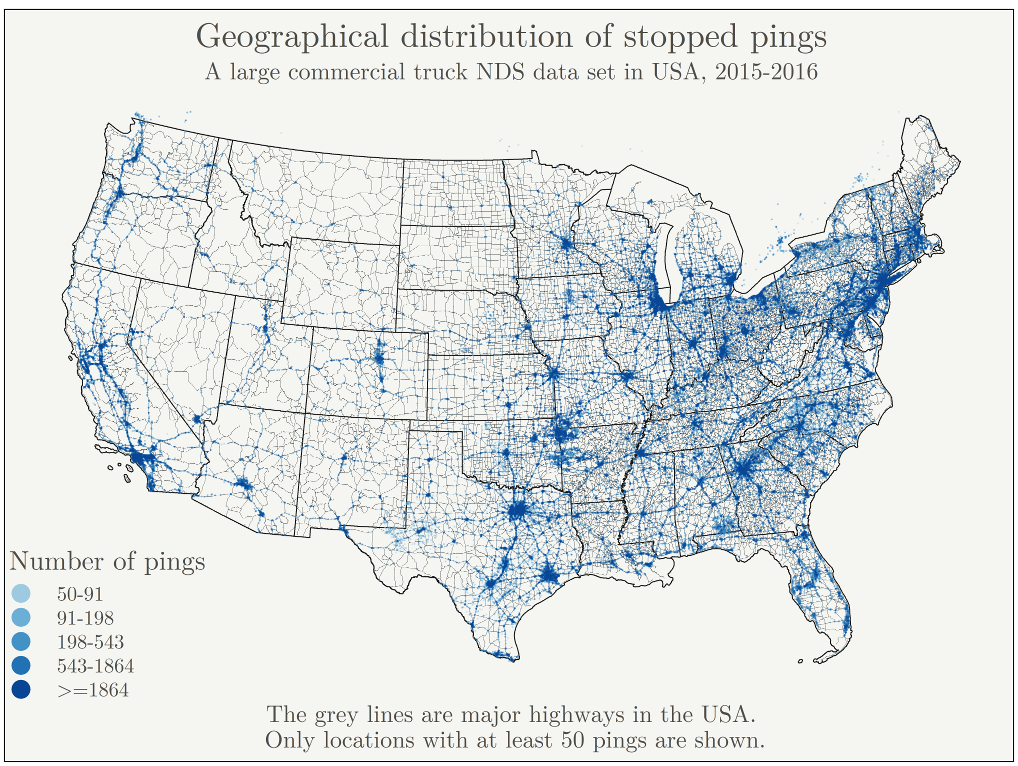
In addition to the sensor-based data, the trucking company provided a crash dataset, which included the anonymized driver ID, state, city, report time, the number fatalities, and the number of injuries. Every row in this dataset corresponded to a unique crash. In total, the 31,828 drivers included in this study were involved in 34,884 crashes, 239 injuries, and 22 fatalities.

2.4. *Data aggregation*

To draw inferences from the ping data, it should be aggregated to provide insights about trips, routes, and/or any other unit of analysis of interest. In this study, we aggregated the ping data into trips based on the approach of Pande et al. (2017). Specifically, we define a trip as a continuous period of driving, which can contain sub-periods of stopping as long the length of any period is under 30 minutes. In our analysis, we sorted the original ping by driver using the observed timestamp for each driver. Then, if the ping speed



(a) Active pings captured from the 31,828 truck drivers from April 1, 2015 to March 31, 2016.



(b) Inactive pings captured from the 31,828 truck drivers from April 1, 2015 to March 31, 2016.

Figure 1: Geographical point patterns of active and inactive pings in a large commercial NDS data set.

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.

The aggregation of pings into trips is demonstrated in Figure 2, where the x-axis shows the date and time of pings, and the y-axis captures the speed (in miles per hour, MPH). Each point represented a ping at that date and time, 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.

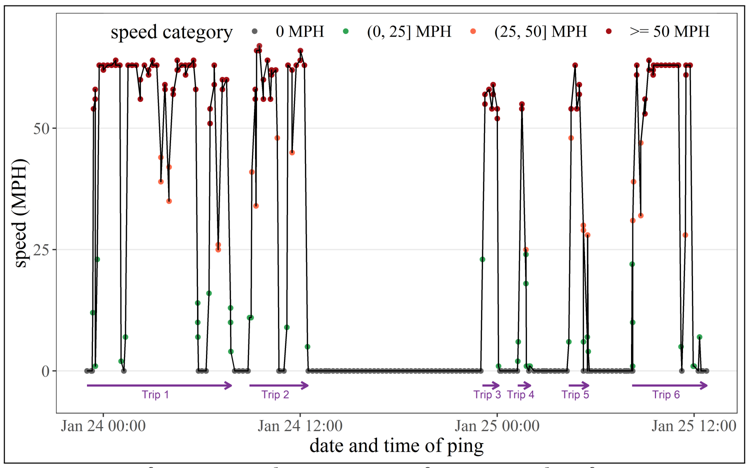


Figure 2: Data aggregation from ping data to trips for a sample of a commercial NDS data set.

The traveled distance within a trip was computed using a three sequential-step procedure. First, we sorted the ping data by driver, date, and time. Then, in step 2, we computed the distance between consecutive ping locations for each driver using the haversine method (Sinnott, 1984), which assumes a spherical earth and ignores all ellipsoidal effects. Although this haversine method does not reflect the true driving distance since it calculates the shortest distance between two points, it is the most feasible algorithm we could use for this huge data set. The third step involved summing up all the distances traveled within a trip for each driver. This procedure 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 total number of miles recorded is consistent with the estimates provided by the company pertaining to their total annual incurred miles. Furthermore, the median trip time was 2.61 hours and the median number of miles per trip was 77.06 miles. A summary of the total number of events described in Sections 2.1-2.4 is provided in Table 3.

3. Methodology

3.1. Data preparation for statistical modeling

With the modeling objectives in mind, we can categorize the data from Section 2 into three statistical categories. First, our outcome/response variables capture the number of (a) crashes, (b) injuries, and (c)

Table 3: A summary of the trucking data that is analyzed in this study.

Pings		Trips		SCEs		Crashes	
Total pings	1,494,678,173	Total trips	18,740,142	Total SCEs	450,758	Crashes	34,884
Drivers	31,828	Total miles	2,320,967,467	Headway	170,421	Injuries	239
		Total hours	65,646,731	Hard brakes	218,419	Fatalities	22
		Median miles	77.06	Collision mitigation	55,243		
		Median hours	2.61	Rolling stability	6,675		

fatalities reported for a given driver. Thus, the outcome variables capture data that is aggregated/added over the 1-year period of data collection. Note that we hypothesize that these three outcome variables should be positively correlated with the total miles driven, and hence we also record the total miles driven by each driver. The second category includes our primary predictor variables of interest, where the rate of SCE per 10,000 miles is computed for each SCE, i.e., four rates are computed for each driver capturing headway, hard brakes, rolling stability and forward collision mitigation events. We also compute a pooled SCE rate per 10,000 miles for each driver, representing the aggregation of the four SCEs (i.e. the rate if any of the four SCEs were observed). Note that all five SCE predictor variables are also based on the driver's total driving activity over the course of the data collection period. The third, and final category, captures the following driver-specific covariates: (a) age, (b) gender, (c) mean ping speed, (d) business unit, and (e) driver type. With the exception of the computed mean ping speed, the other four covariates are obtained using the company provided demographics table. A summary of the computed variables is provided in Table 4. Note that the distribution of the other covariates was provided in Table 2.

Table 4: A summary of the computed response, predictor variables and covariate.

Variable	Mean ± SD
Number of crashes per driver	1.10 ± 1.69
Headways per driver per 10,000 miles	5.35 ± 27.01
Hard brakes per driver per 10,000 miles	6.86 ± 22.75
Rolling stability initiations per driver per 10,000 miles	0.21 ± 1.20
Collision mitigation initiations per driver per 10,000 miles	1.74 ± 4.66
Ping speed per driver	29.85 ± 7.23

The final step in preparing the data for modeling was to examine whether the number of predictors and/or covariates can be reduced based on a correlation analysis. Note that the variables included both continuous (primary predictors, age and mean ping speed) and polytomous (gender, business unit, and driver type) variables. To account for the mixed variable types, we adopted the approach of Revelle et al. (2010, 2016) to compute the Pearson, polychoric and polyserial correlation coefficients for the pairwise evaluation of continuous, polytomous and mixed variables, respectively. We present the results in Figure 3.

Based on Figure 3, there are two observations to be highlighted. First, the Pearson correlation coefficient values between any two rates of SCEs were small (< 0.2). Second, none of the correlation coefficients exceeded 0.6 in magnitude. Accordingly, we concluded that the resulting statistical models will not have

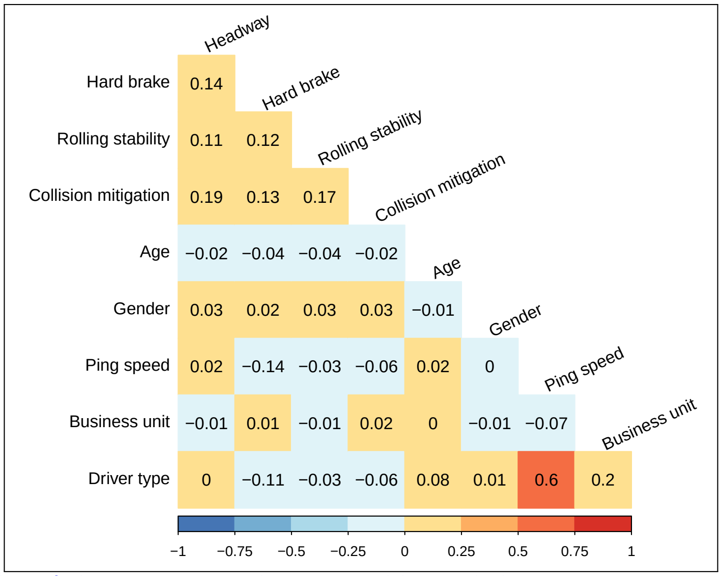


Figure 3: A correlation plot of the primary predictor variables and the covariates. Note that multiple correlation coefficients were used to account for the mixed (i.e., categorical and numeric) nature of the variables. The top four variables are the SCEs and the others are the predictor variables.

any serious multicollinearity issues (which will be examined in greater detail during the model assessment stage).

3.2. *Modeling strategy and its relation to the examined research questions*

Note that the outcome variable (number of crashes, injuries, or fatalities) is a strictly non-negative integer. In the literature, Poisson regression or negative binomial models have been commonly applied for this type of analysis (e.g., see Table 1). By design, the Poisson regression model assumes equal mean and variance of outcome distribution, which can lead to inferior performance in the cases of potential overdispersion or underdispersion issues in the data (Lord and Mannering, 2010). Negative binomial modeling can adjust for the variance independently from its mean. Consequently, we employ the latter approach. We also adopted Bayesian rather than frequentist estimation approach since it provides more flexibility in specifying statistical models when compared to traditional maximum likelihood estimation methods (Dunson, 2001). Furthermore, in the case of rare-events, even relatively flat priors can increase the precision of parameter estimation. Using the Bayesian negative binomial models, we can investigate the three research questions introduced in Section 1.2. Table 5 shows the different combinations of samples and models.

3.3. *Statistical models*

3.3.1. *Model specification*

Let Y_i denote an outcome variable (i.e., the number of crashes, injuries or fatalities) over a distance of T_i miles for the ith driver. Each of the three outcomes was modeled in different Bayesian negative binomial models. We assume that Y_i has a negative binomial distribution with mean parameter μ_i and a common

Table 5: An overview of how the data is sampled and modeled to examine the three research questions.

Research question	**Samples**	**Model**
Question 1: Are SCEs associated with crashes, injuries, and fatalities?	All data	# Crashes = f(Pooled SCE rate and all covariates) # Crashes = f(Individual rate for each of the 4 SCEs and all covariates) # Injuries = f(Pooled SCE rate and all covariates) # Injuries = f(Individual rate for each of the 4 SCEs and all covariates) # Fatalities = f(Pooled SCE rate and all covariates) # Fatalities = f(Individual rate for each of the 4 SCEs and all covariates)
Question 2: Is the association between crashes and SCEs consistent for all four different SCEs?	All data	# Crashes = f(Headway rate and all covariates) # Crashes = f(Hard brake rate and all covariates) # Crashes = f(Rolling stability initiation rate and all covariates) # Crashes = f(Collision mitigation rate and all covariates)
Question 3: Does the association between crashes and SCEs hold for the different business units and driver types?	Stratified sampling by: Dedicated ∩ Local Dedicated ∩ Regional Dedicated ∩ Over-the-Road Intermodal ∩ Local Intermodal ∩ Regional Final Mile ∩ Regional Final Mile ∩ Over-the-Road	# Crashes = f(Individual rate for each of the 4 SCEs, mean speed, age and gender). Note that the model above is applied to each individual stratified sample.

auxiliary parameter ϕ. The corresponding probability mass function of Y_i is parameterized as

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

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 an additive effect on the logarithm of rate of crashes μ_i, we have the following log-linear Bayesian negative binomial regression:

$$\begin{aligned} Y_i &\sim \text{Negative Binomial}(T_i \times \mu_i,\ \phi) \\ \log \mu_i &= \alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK}, \end{aligned} \tag{2}$$

where the total miles driven T_i is considered as an offset variable to account for the mileage difference among drivers. Also, α_0 is the intercept and $\beta_j,\ j = 1, 2, \ldots, J$ is the coefficient of the j-th primary predictor x_{ij}. $J = 1$ for pooled SCE rate models (x_{i1} represents the rate of any SCE) and one SCE rate at a time models (x_{i1} represents the rate of headway, hard brakes, rolling stability, or collision mitigation, with only one of them used at a time). Note that the values of J and K depend on the application of the models as shown in Table 5. Specifically, $J = 4$ for four SCEs models ($x_{i1}, x_{i2}, x_{i3}, x_{i4}$ represent the rate of headway, hard brakes, rolling stability, and collision mitigation, respectively, with all four variables used in the same model at a time). The parameter $\theta_k,\ k = 1, 2, \ldots, K$, is the coefficient of the k-th covariate z_{ik}, and $K = 4$ (age, mean speed, gender) for models stratified by business units and driver types, while $K = 8$ (age, mean speed, gender, business unit, and driver types) for non-stratified models. Gender, business units, and driver types were coded using one-hot encoding: (a) gender included female (reference group), male, and unknown; (b) business units included dedicated (reference group), intermodal, and final-mile; and (c) driver types included local (reference group), regional, and over-the-road. We took relatively noninformative priors for β_k and ϕ. Specifically, we assumed $\alpha_0,\ \beta_1, \ldots, \beta_J,\ \theta_1, \ldots, \theta_K \sim \text{Normal}(0, 10^2)$ and $\phi \sim \text{Exponential}(1)$.

3.3.2. Bayesian estimation and parameter interpretation

We applied the Hamiltonian Monte Carlo procedure to estimate the posterior distributions of all parameters. Compared to its two predecessor Markov chain Monte Carlo samplers (the Metropolis-Hastings algorithm and Gibbs sampling), Hamiltonian Monte Carlo is much more efficient in making valid proposal samples, and there exists a well-developed statistical package implementing this algorithm (McElreath, 2020). To make sure the Hamiltonian Monte Carlo converged to the true posterior distributions, 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).

In the Bayesian setting, parameters are considered as random variables that have probabilistic distributions instead of unknown fixed values, so no p-values can be reported here. The posterior mean and 95% credible intervals (CIs) of the incidence rate ratios ($\exp(\beta)$) are reported instead. The interpretation of the incidence rate ratios in this Bayesian negative binomial model is as follows: as the number of SCEs per 10,000 miles increases by one unit, the number of crashes per mile is multiplied by $\exp(\beta)$. A 95% credible interval is the interval such that is the posterior probability of the parameter of interest falling within that range given the data is 95% (McElreath, 2020). If the 95% CI of the incidence rate ratio includes one, then one is a plausible value for the true incidence rate ratio in this case, and the parameter of the variable will be deemed statistically insignificant. If the 95% credible interval excludes one, then the parameter will be considered as statistically significant.

3.3.3. Model validation and assessment

Since four different SCEs and/or multiple covariates were included in the models at the same time, it is important to check for the presence of multicollinearity. Recall that in Section 3.1, we examined the correlations among pairs of predictors and found that the pairwise correlations are small. Here, we attempt to investigate whether a linear dependence exists among three or more of our variables through computing the variance inflation factors. If the regressors are uncorrelated, the variance inflation factor obtains its minimum value of 1. In statistical practice, a inflation factor less than 4 requires no additional investigation of linear dependence among the regressors and values greater than 10 indicate serious multicolinearity requiring model corrections (Vatcheva et al., 2016).

We used Pareto smoothed importance-sampling leave-one-out 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, this algorithm uses fast, efficient, and stable importance sampling weights to approximate leave-one-out cross-validation (Gelfand et al., 1992; Gilks et al., 1996). It is used to estimate the expected log predicted density (ELPD_LOO), the effective number of parameters (P_LOO), and the LOO Information Criterion (LOOIC) for a model. The ELPD_LOO is the average leave-one-out predictive density given the data (similar to the log likelihood in maximum likelihood estimation), and larger ELPD_LOO values indicate better model prediction. LOOIC is an information criterion similar to the widely applicable information criterion: its value is close to $-2\times$ELPD_LOO, and smaller LOOIC suggest better models (Vehtari et al., 2017). Compared with other statistics such as the widely applicable information criterion, deviance information criterion, and other variants (Spiegelhalter et al., 2002; Watanabe, 2010), Pareto smoothed importance-sampling leave-one-out is both fast and stable in computing. P_LOO is an estimate of the effective number of parameters, and the value should be close to the real number of parameters in a well-specified model. These statistics (ELPD_LOO, P_LOO, and LOOIC) are calculated to compare the model predictive performance and identify any potential badly-specified models. 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 Section 4.

3.4. Statistical software used

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 leave-one-out cross-validation statistics were computed using the `loo()` function in `loo` package (Vehtari et al., 2019). To facilitate the adoption of our methodology, we provide a link of a compiled **R** Markdown file, containing the code for data aggregation and model building, and their associated results, in the supplementary materials section.

4. Results

4.1. The association of SCEs with crashes, injuries, and fatalities among commercial truck drivers

Table 6 presents the Bayesian negative binomial models' results explaining the variation in number of crashes, injuries, and fatalities separately using either a pooled SCE predictor (i.e., four SCEs as one variable), or four SCE predictors (i.e., the four SCEs represented using four variables). In column 2, we present the results for explaining the differences in the number of crashes accumulated by each driver as a function of pooling the counts of SCEs, while accounting for differences in the observed covariates and differences in total miles driven. The coefficient of the "all SCEs" predictor in the second column suggests that a unit increase in the number of any type of SCE per 10,000 miles was associated with an 8.4% (95% CI: 8.0-8.8%) increase in the rate of crashes. When broken down by the four different SCEs in the third column, one can note the following: (a) the coefficients of all four predictors and their 95% CIs are larger than 1, indicating (similar to the pooled model) that an increase in any of the four SCEs is associated with an increase in the number of crashes; (b) a unit increase in the number of instances of rolling stability system initiation was associated with the largest, 50.4% (95% CI: 41.4-60.0%), in the number of crashes per mile; and (c) when holding other SCEs constant, a unit increase in either the initiation of the rolling stability or collision mitigation systems has larger effects when compared to increases in hard brakes or headyway alerts. An implication from observation (c) is that the effect of driver-less maneuvers are larger than alerts and actions where the driver may be involved. Conceptually, this is an intuitive result since these maneuvers can be categorized as more "aggressive last-minute" interventions that attempt to mitigate a crash by overriding the driver's control. The coefficients of the covariates suggested that older drivers, higher average speed, male, as well as final-mile and intermodal (compared to dedicated) drivers, were associated with lower rate of crashes.

Compared with the models for crashes, the results for injuries and fatalities (columns 4-7) tend to be less conclusive since the number of recorded injuries and fatalities are much smaller. In the injuries-pooled model (column 4), a 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 (statistically insignificant) 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 (columns 6 and 7), all 95% CIs of incidence rate ratios included one and the CIs were very wide.

4.2. The consistency of the association between crashes and the four different SCEs

Table 7 shows the estimates of posterior incidence rate ratios and their CIs in the Bayesian negative binomial models for all the included drivers. Note that we include both the pooled and four SCE models from Table 6 to facilitate the comparison with the four different models corresponding to when each of the headways, hard brakes, rolling stability or collision mitigation predictors were included. From the presented results, similar conclusions can be made to the insights gained from examining the pooled and four SCE models. Specifically, the coefficients of each of the main predictors were larger than one (with the associated 95% credible interval excluding one), providing statistically strong evidence that the rates of the individual SCEs were positively associated with the rates of crashes. Furthermore, the rolling stability coefficient

Table 6: Bayesian negative binomial regressions with the rate of SCEs predicting crashes, injuries, and fatalities

Variables	Crashes: pooled	Crashes: four SCEs	Injuries: pooled	Injuries: four SCEs	Fatalities: pooled	Fatalities: four SCEs
Intercept	0.054 (0.047, 0.062)	0.048 (0.042, 0.054)	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	1.084 (1.080, 1.088)		1.087 (1.048, 1.136)		0.973 (0.791, 1.149)	
Headways		1.033 (1.026, 1.040)		1.061 (0.961, 1.181)		0.955 (0.592, 1.478)
Hard brakes		1.081 (1.075, 1.087)		1.080 (0.995, 1.177)		0.957 (0.652, 1.387)
Rolling stability		1.504 (1.414, 1.600)		1.773 (0.684, 5.439)		1.631 (0.043, 102.8)
Collision mitigation		1.222 (1.198, 1.245)		1.174 (0.987, 1.535)		0.866 (0.200, 3.632)
Age	0.992 (0.990, 0.993)	0.992 (0.991, 0.993)	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	0.979 (0.976, 0.982)	0.982 (0.979, 0.985)	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	0.817 (0.756, 0.886)	0.808 (0.754, 0.867)	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	0.975 (0.785, 1.199)	0.954 (0.777, 1.149)	1.022 (0.094, 8.499)	0.993 (0.092, 9.338)	0.093 (0, 76.2)	0.093 (0, 115.6)
Business unit: Inter-modal	0.698 (0.670, 0.727)	0.717 (0.690, 0.745)	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	0.907 (0.861, 0.954)	0.897 (0.852, 0.943)	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	1.071 (0.994, 1.151)	1.094 (1.022, 1.174)	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	1.003 (0.957, 1.045)	1.012 (0.969, 1.057)	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	31828	31828
ELPD_LOO	-39985.2 (236.5)	-39770.2 (233.5)	-1134.5 (80.8)	-1137.3 (81.1)	-182.4 (37.9)	-182.4 (37.9)
P_LOO	18.1 (1.1)	30 (2.4)	13.9 (3.6)	16.4 (4)	11.3 (3.2)	11.3 (3.2)
LOOIC	79970.4 (472.9)	79540.5 (467.1)	2269.1 (161.5)	2274.6 (162.1)	364.7 (75.7)	364.7 (75.7)

Notes:
The SCEs were measured as the number of events per 10,000 miles driven.
Incidence rate ratios and their associated 95% credible intervals are reported for all variables (predictors and covariates).
For the fit statistics, (.) indicates the standard error of the computed statistic.

is the largest, followed by the crash mitigation coefficient, confirming that an increase in more aggressive interventions results in more crashes.

4.3. The influence of business units and driver types on the association between crashes and SCEs

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 8. The posterior incidence rate ratios and CIs of four SCEs are consistent with those in Table 6. 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 and 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.

4.4. Diagnostic statistics and model selection

We computed the variance inflation factors described in Section 3.3.3 to test for multicollinearity. Table 9 captures the results for all the negative binomial models where individual predictors are used for each of the SCEs. The results capture all three investigated outcomes (crashes, fatalities and injuries) as well as the stratification of the drivers' data set by business unit and driver type. From the table, all variance inflation factors are less than or equal to 1.3. Hence, we can conclude that the variation of each of the regression coefficients are not inflated and that there is no evidence of any multicollinearity issues (Vatcheva et al., 2016).

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 6, 7, 8), were similar to the total number of parameters in the models. These two results suggest that the 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. For example, in Table 7, the "Four SCEs" model has the lowest LOOIC (79,540.5) among the six models. However, the standard errors of the LOOIC statistic (in parentheses) suggest that the model with all four SCEs was not significantly better than the pooled model.

To investigate the models' predictive accuracy, we adopt the approach of Gelman et al. (2013, Section 6.3) who suggested simulating some function of the data and parameter, and comparing it with the observed value of a particular quantity. For our trucking safety application, we examined the proportion of zero crashes since it corresponds to a crash-free trip, which is of interest to truck drivers and operators

Table 7: Bayesian negative binomial regressions predicting crashes with different combination of SCEs

Variables	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)
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)
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)

Notes:
The SCEs were measured as the number of events per 10,000 miles driven.
Incidence rate ratios and their associated 95% credible intervals are reported for all variables (predictors and covariates).
For the fit statistics, (.) indicates the standard error of the computed statistic.

Table 8: Bayesian negative binomial regressions with SCEs predicting crashes, stratified by business units and driver types

variables	Dedicated Local	Dedicated Over-the-road	Dedicated Regional	Inter-modal Local	Inter-modal Regional	Final-mile Over-the-road	Final-mile 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)

Notes:
The SCEs were measured as the number of events per 10,000 miles driven.
Incidence rate ratios and their associated 95% credible intervals are reported for all variables (predictors and covariates).
For the fit statistics, (.) indicates the standard error of the computed statistic.

alike. The probability of having zero crashes is, of course, an unknown quantity, but its posterior distribution can be estimated by simulating samples using Hamiltonian Monte Carlo. In this section, we limit our analysis to the models whose outcomes were crashes since the accident and fatality models indicated that our observed events were insufficient for statistical inference (based on the size of the credible interval in Table 6).

Figure 4 shows the posterior distributions, which are indicated by the histograms in light blue, for the posterior probability of zero crashes under each of the six models considered in Table 7. The observed

Table 9: Variance inflation factor test for multicollinearity.

Outcome	Crash	Fatality	Injury	Crash	Crash	Crash	Crash	Crash	Crash	Crash
Samples	All drivers	All drivers	All Drivers	Dedicated ∩ Local	Dedicated ∩ Regional	Dedicated ∩ OTR	Intermodal ∩ Local	Intermodal ∩ Regional	Final-mile ∩ Regional	Final-mile ∩ OTR
Headways	1.030	1.030	1.030	1.049	1.024	1.012	1.044	1.040	1.073	1.037
Hard brakes	1.031	1.031	1.031	1.027	1.063	1.047	1.020	1.031	1.077	1.184
Rolling stability	1.024	1.024	1.024	1.029	1.013	1.047	1.024	1.010	1.053	1.159
Collision mitigation	1.038	1.038	1.038	1.024	1.048	1.056	1.056	1.046	1.118	1.236
Age	1.006	1.006	1.006	1.003	1.007	1.005	1.004	1.004	1.004	1.011
Ping speed	1.270	1.270	1.270	1.007	1.012	1.012	1.015	1.015	1.011	1.047
Gender	1.005	1.005	1.005	1.001	1.002	1.005	1.004	1.004	1.010	1.008
Business unit	1.096	1.096	1.096							
Driver type	1.140	1.140	1.140							

proportion of zero crashes is indicated by the vertical line in each part of Figure 4. For all six models, the observed proportion of zero crashes was considerably less than what would be predicted by the model. Note that the magnitude of this prediction bias is small, usually around 0.015. In other words, while both models (with and without business units and driver types) perform reasonably well in predicting the mean numbers of SCEs, the model with business units and driver types does a better job predicting the proportion of zero crashes. This suggests that different business units and driver types should be accounted for in the model.

Based partly on the result from Figure 4, we ran the model with all four SCEs (model 2) separately for each of the seven business units and driver types. The corresponding posterior predictive check for zero crashes is shown in Figure 5. Here, the vertical lines are much closer to the simulated posterior distribution. This suggests that different business units and driver types should be accounted for in the model.

5. Discussion and conclusions

5.1. Summary of the main contributions

In line with previous studies on the association between crashes and SCEs in NDSs (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 a positive association between SCEs and crashes based on routinely collected NDS and kinematic data

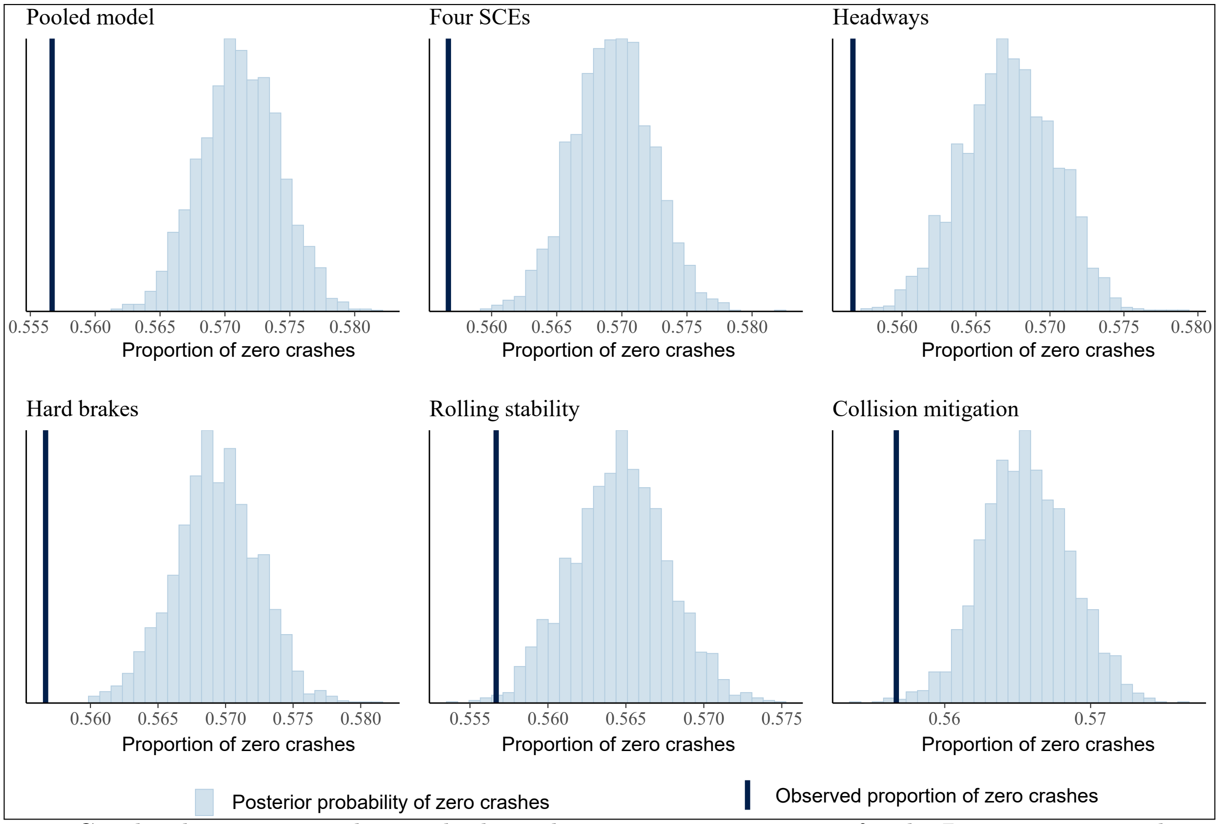


Figure 4: Graphical posterior predictive checks with zero count test statistic for the Bayesian negative binomial models for all drivers. The x-axis is the proportion of zero crashes and y-axis is probability density. The solid black line is the observed proportion, while the light blue histogram is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

from 31,828 American truck drivers. 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 for different types of SCEs: 3.3% (95% CI: 2.6-4.0%) for headways, 8.1% (95% CI: 7.5-8.7%) for hard brakes, 50.4% (95% CI: 41.4-60.0%) 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. Furthermore, this study demonstrates that the "severity" of the SCE is associated with the crash rate, where the two automated maneuvers (involving the initiation of the forward-collision mitigation and rolling stability systems) were shown to have a statistically significant larger effects on crash rates when compared to hard brakes and head way (which can be seen as less severe maneuvers/alerts).

The current study contributes to the existing literature in three respects. First, this paper overcomes the small sample size issues in previous crashes and crash surrogates papers, which typically includes 300 or fewer drivers or vehicles and fewer than 100 crashes (Guo et al., 2010; Gitelman et al., 2018). Our study involves 1,000 times as many driving hours and miles and includes more than $30,000$ commercial truck drivers and $30,000$ crashes; this allows us to investigate the association between four different types of

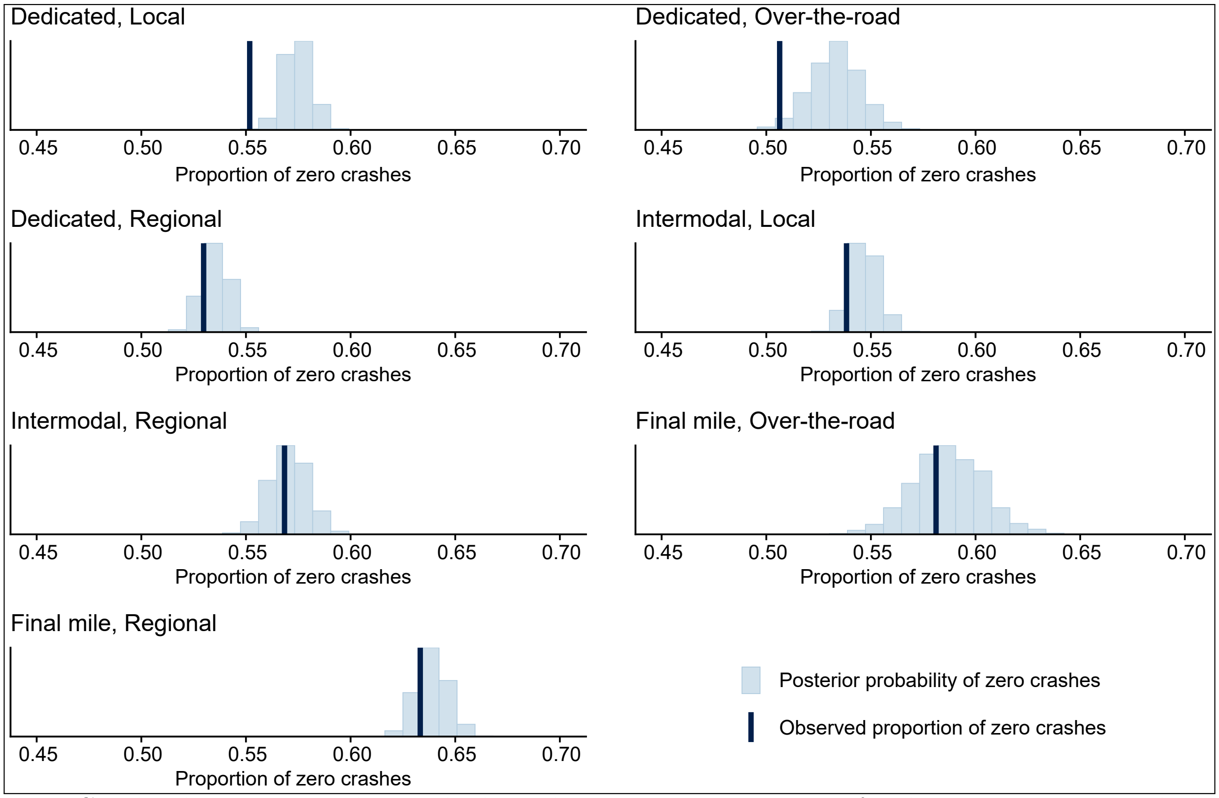


Figure 5: Graphical posterior predictive checks with zero count test statistic for the Bayesian negative binomial models, stratified by business unit and driver types. The x-axis is the proportion of zero crashes and y-axis is probability density. The solid black line is the observed proportion, while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

SCEs and crashes, as well as stratified analyses across business units and driver types. Second, the evidence of the association between crashes and crash surrogates among truck drivers has been scarce. Our study gives insights to this less studied field using a nationwide large-scale sample. Third, this paper explores the association between SCEs and human injuries and fatalities, which has not been investigated in previous papers but represents important research questions that require detailed study given that they constitute an important component of truck routing models used in practice (Hu et al., 2020).

5.2. Practical relevance of our work to trucking operators

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 little evidence has been provided. Our study is the only study, to the best of our knowledge, that quantifies this association for commercial truck drivers. Consequently, our results allow for the use of SCEs (specifically the initiation of the rolling stability system, the initiation of the collision mitigation system, hard brakes and/or headway alerts) as proxies for crashes. The ability to use a more frequent outcome will allow for more statistical power when investigating variables that affect (trucking) safety.

From a trucking operator's perspective, the SCE data used in this study are routinely collected with the wide-scale adoption of sensor-based monitoring systems in practice. In many cases, the data have not been used in enhancing/transforming safety operations because: (a) it is unclear whether or how the SCEs and crashes are associated; (b) the trucking operators may not know how to aggregate the acquired ping data into a more meaningful unit of analysis; (c) the modeling/coding of this data may be difficult especially for smaller operators with limited statistical and engineering resources. In this study, we have demonstrated the positive relationship of the SCEs with the number of crashes recorded for a large number of drivers. Furthermore, our stratified models by driver type and business unit indicate that the incorporation of such information can enhance the predictions obtained from the model. To facilitate the adoption of our work by industry practitioners, we host a compiled **R** Markdown document containing our code, results and analysis on a freely accessible website (see link in the supplementary materials section). The use of the open-source **R** programming language can make the adoption of our work more accessible for trucking operators of different sizes.

The knowledge of the positive association between the four SCEs and crashes can be incorporated by trucking operators using several approaches. First, recent statistics indicate that more than 90% of traffic crashes are influenced by driver behavior (Federal Highway Administration, 2019). While our naturalistic driving data does not include video images, it explicitly captures important behavioral factors such as driving speed, aggressive driving through headway alerts, and potential distraction/drowsiness with an increased rate of the three other SCEs (at least when compared to drivers on similar routes). Thus, trucking operators can use our estimated model coefficients, e.g., 50.4% for the initiation of the rolling stability system, in driver training/education. Second, operators can provide incentives to their drivers to reduce their recorded number of SCEs through behavioral based safety programs (Jun et al., 2007). Third, by examining the operator's historical record of SCE data, operators can develop scheduling and routing policies that attempt to minimize the number of recorded SCEs (Mehdizadeh et al., 2020; Hu et al., 2020).

5.3. Limitations and suggestions for future research

This study has several limitations that can be investigated in future studies. 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 (Guo et al., 2010). 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 in either of the two models is not strong given the small number of fatalities. Third, the data does not include traffic or weather variables, which are important predictors of crashes. Fourth, this study used a cross-sectional observatory design and no experiments were involved, so 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. Fifth, the haversine method to calculate the distance may not reflect the true miles driven by the sample drivers since it calculates the shortest distance between two points. Despite these facts, it is the most feasible and cost-effective algorithm to estimate the distance between two points for this study.

Despite these limitations, we demonstrate in this paper the positive association between safety critical events and crashes/injuries/fatalities. Our study features several orders of magnitude more data volume than previous studies in terms of number of drivers, hours driven, miles driven, number of crashes, and the count of surrogate events, as well as a previously unexamined population: commercial truck drivers. These results indicate that the emerging widespread use of technology for monitoring kinematic events can effectively be used for naturalistic driving studies.

Supplemental materials

To encourage future research and/or adoption of our work by trucking operators, we provide an **R** Markdown document which provides the code used for aggregating the ping data into trips and statistical modeling. Our code and analysis can be freely accessed at https://caimiao0714.github.io/Github-SCE-crash/. We provide a sample CSV file, where we masked the driver ID as well as the coordinates (by rounding them to one digit) at https://github.com/caimiao0714/Github-SCE-crash/blob/master/data/sample_ping.csv. The purpose of providing the sample file is to provide some insight into the ping data frequency, shape, and speed values. Note that we cannot provide the ping or the aggregated data based on the terms of our nondisclosure agreement.

Acknowledgments

This work was supported in part by the National Science Foundation (CMMI-1635927 and CMMI-1634992); the Ohio Supercomputer Center (PMIU0138 and PMIU0162); the University of Cincinnati Education and Research Center Pilot Research Project Training Program; the Transportation Informatics Tier I University Transportation Center (TransInfo); and a Google Cloud research grant.

References

Bendix®, 2007. Bendix® ABS-6 Advanced with ESP® Stability System - frequently asked questions to help you make an intelligent investment in stability. Bendix Commercial Vehicle Systems LLC, a member of the Knorr-Bremse Group. https://www.bendix.com/media/documents/products_1/absstability/truckstractors/StabilityFAQ.pdf. [Published March 2007; accessed April 19, 2020].

Blower, D., 2017. Estimating motor carrier management information system crash file underreporting from carrier records. Federal Motor Carrier Safety Administration. Office of Analysis, Research, and Technology. FMCSA-RRR-16-025. https://doi.org/10.21949/1502548. [Published online August 1, 2017; accessed April 19, 2020].

Cooper, P., 1984. Experience with traffic conflicts in Canada with emphasis on "post encroachment time" techniques, in: International calibration study of traffic conflict techniques. Springer, pp. 75–96.

Dingus, T.A., Hanowski, R.J., Klauer, S.G., 2011. Estimating crash risk. Ergonomics in Design 19, 8–12.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., et al., 2006. 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. URL: https://trid.trb.org/view/783477.

Dunson, D.B., 2001. Commentary: practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology 153, 1222–1226.

Eenink, R., Barnard, Y., Baumann, M., Augros, X., Utesch, F., 2014. UDRIVE: the European naturalistic driving study, in: Proceedings of Transport Research Arena, TRA 2014, 14-17 Apr 2014, Paris, France. IFSTTAR. URL: http://eprints.whiterose.ac.uk/93078/.

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

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

Federal Highway Administration, 2019. Human factors. U.S. Department of Transportation, https://highways.dot.gov/research/research-programs/safety/human-factors. [Updated December 2, 2019; accessed April 29, 2020].

FMCSA, 2019. Large Truck and Bus Crash Facts 2017. https://www.fmcsa.dot.gov/safety/data-and-statistics/large-truck-and-bus-crash-facts-2017. [Online; accessed 20-April-2020].

Gelfand, A.E., Dey, D.K., Chang, H., 1992. Model determination using predictive distributions with implementation via sampling-based methods. Technical Report. Department of Statistics, Stanford University. URL: https://statistics.stanford.edu/sites/g/files/sbiybj6031/f/SOL%20ONR%20462.pdf.

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

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

Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Model determination using sampling-based methods. Chapman and Hall, London. chapter 9. pp. 145–161.

Gitelman, V., Bekhor, S., Doveh, E., Pesahov, F., Carmel, R., Morik, S., 2018. 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.

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

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

Grove, K., Atwood, J., Hill, P., Fitch, G., DiFonzo, A., Marchese, M., Blanco, M., 2015. Commercial motor vehicle driver performance with adaptive cruise control in adverse weather. Procedia Manufacturing 3, 2777–2783.

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

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

Hankey, J.M., Perez, M.A., McClafferty, J.A., 2016. Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets. Technical Report. Virginia Tech Transportation Institute. URL: http://hdl.handle.net/10919/70850.

He, Z., Hu, J., Park, B.B., Levin, M.W., 2019. Vehicle sensor data-based transportation research: Modeling, analysis, and management. Journal of Intelligent Transportation Systems 23, 99–102.

Heinrich, H.W., 1931. Industrial Accident Prevention: A Scientific Approach. New York & London: McGraw-Hill Book Company, Inc.

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

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

Hu, Q., Cai, M., Mohabbati-Kalejahi, N., Mehdizadeh, A., Yazdi, A., Ali, M., Vinel, A., Rigdon, S.E., Davis, K.C., Megahed, F.M., 2020. A review of data analytic applications in road traffic safety. part 2: Prescriptive modeling. Sensors 20, 1096. doi:10.3390/s20041096.

Hydén, C., 1987. The development of a method for traffic safety evaluation: The Swedish traffic conflicts technique. Bulletin Lund Institute of Technology, Department .

Imprialou, M., Quddus, M., 2019. Crash data quality for road safety research: current state and future directions. Accident Analysis & Prevention 130, 84–90.

John, S., 2019. 11 incredible facts about the $700 billion US trucking industry. Business Insider: Markets Insider. https://markets.businessinsider.com/news/stocks/trucking-industry-facts-us-truckers-2019-5-1028248577. [Published online June 3, 2019; accessed April 19, 2020].

Jun, J., Ogle, J., Guensler, R., 2007. Relationships between crash involvement and temporal-spatial driving behavior activity patterns: use of data for vehicles with global positioning systems. Transportation Research Record 2019, 246–255.

Katrakazas, C., Quddus, M., Chen, W.H., Deka, L., 2015. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. Transportation Research Part C: Emerging Technologies 60, 416–442.

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

McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.

Mehdizadeh, A., Cai, M., Hu, Q., Yazdi, A., Ali, M., Mohabbati-Kalejahi, N., Vinel, A., Rigdon, S.E., Davis, K.C., Megahed, F.M., 2020. A review of data analytic applications in road traffic safety. part 1: Descriptive and predictive modeling. Sensors 20, 1107. doi:10.3390/s20041107.

Neale, V.L., Dingus, T.A., Klauer, S.G., Sudweeks, J., Goodman, M., 2005. An overview of the 100-car naturalistic study and findings, in: Proceedings - 19th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Washington, D.C., June 6-9, 2005, National Highway Traffic Safety Administration. p. 10p. URL: https://trid.trb.org/view/815278.

NHTSA's National Center for Statistics and Analysis, 2019. 2017 data: large trucks. U.S. Department of Transportation. National Highway Traffic Safety Administration. Traffic Safety Facts. DOT HS 812 663. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812663. [Published online January 2019; accessed April 19, 2020].

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

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

Revelle, W., Condon, D.M., Wilt, J., French, J.A., Brown, A., Elleman, L.G., 2016. Web and phone based data collection using planned missing designs. Sage Publications, Inc. pp. 578–595. doi:10.4135/9781473957992.

Revelle, W., Wilt, J., Rosenthal, A., 2010. Individual Differences in Cognition: New Methods for Examining the Personality-Cognition Link. Springer New York, New York, NY. pp. 27–49. doi:10.1007/978-1-4419-1210-7_2.

Risser, R., 1985. Behavior in traffic conflict situations. Accident Analysis & Prevention 17, 179–197.

Saleh, J.H., Saltmarsh, E.A., Favaro, F.M., Brevault, L., 2013. 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.

Shi, Q., Abdel-Aty, M., 2015. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transportation Research Part C: Emerging Technologies 58, 380–394.

Shmueli, G., et al., 2010. To explain or to predict? Statistical Science 25, 289–310.

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

Sinnott, R.W., 1984. Virtues of the haversine. Sky and Telescope 68, 158–159.

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

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

The White House, 2020. Remarks by President Trump celebrating America's truckers. https://www.whitehouse.gov/briefings-statements/remarks-president-trump-celebrating-americas-truckers/. [Issued on April 16, 2020; accessed April 19, 2020].

Theofilatos, A., Yannis, G., Kopelias, P., Papadimitriou, F., 2019. Impact of real-time traffic characteristics on crash occurrence: Preliminary results of the case of rare events. Accident Analysis & Prevention 130, 151–159.

Tsai, Y.T., Alhwiti, T., Swartz, S.M., Megahed, F.M., 2015. The effects of socio-economic and public policy factors on us highway safety. Journal of Transportation Law, Logistics, and Policy 82, 31–48.

Tsai, Y.T., Swartz, S.M., Megahed, F.M., 2018. Estimating the relative efficiency of highway safety investments on commercial transportation. Transportation Journal 57, 193–218.

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

Vatcheva, K.P., Lee, M., McCormick, J.B., Rahbar, M.H., 2016. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale, Calif.) 6.

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

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27, 1413–1432.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., Gabry, J., 2015. Pareto smoothed importance sampling. `arXiv:1507.02646`.

Wali, B., Khattak, A.J., 2020. Harnessing ambient sensing & naturalistic driving systems to understand links between driving volatility and crash propensity in school zones–a generalized hierarchical mixed logit framework. Transportation Research Part C: Emerging Technologies 114, 405–424.

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

Wu, K.F., Aguero-Valverde, J., Jovanis, P.P., 2014. 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.

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

May 31, 2020

Dr. Yafeng Yin
Editor-in-Chief
Transportation Research Part C: Emerging Technologies
Department of Civil & Environmental Engineering
University of Michigan
2350 Hayward Street
Ann Arbor MI 48109-2125, USA

Dear Dr. Yin,

Our paper "The association between crashes and safety-critical events: synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers" has not been submitted to any other journal of any kind. The authors have no conflicts of interest related to the contents of this paper.

On behalf of the research team,

Steven E. Rigdon, Ph.D.
Department of Epidemiology and Biostatistics
College for Public Health and Social Justice
Saint Louis University
3545 Lafayette Avenue
St. Louis, MO 63108, USA