Modeling safety-critical events using trucking naturalistic driving data: A driver-centric hierarchical framework for data analysis

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

[a]*Department of Epidemiology and Biostatistics, Saint Louis University, Saint Louis, MO, 63108, United States*
[b]*Carey Business School, Johns Hopkins University, Baltimore, MD, 21218, United States*
[c]*Department of Industrial and Systems Engineering, Auburn University, Auburn, AL, 36849, United States*
[d]*Department of Computer Science and Software Engineering, Miami University, Oxford, OH, 45056, United States*
[e]*Department of Information Systems and Analytics, Miami University, Oxford, OH, 45056, United States*

Abstract

This study of 497 truck drivers proposes a driver-centric framework. We proposed hierarchical logistic and negative binomial (NB) models with driver-level random intercepts and random slopes for cumulative driving time to predict safety-critical events (SCEs).

Keywords: Trucking, Naturalistic driving studies, Safety-critical events

1. Introduction

The World Health Organization (WHO, 2018) estimated that road injury claimed around 1.4 million lives globally in 2016, which was the eighth leading cause of death. Among all types of vehicles on road, large trucks are a concern since they are more frequently involved in catastrophic crashes. In the United States, National Highway Traffic Safety Administration (2017) reported that 4.3% of registered vehicles were large trucks or buses, but they account for 12.4% of vehicle-related fatalities (Hickman et al., 2018). Truck drivers are often on the road for long routes under on-time demands, complex traffic and weather conditions, with little to no supervision and contact with fellow workers. Therefore, trucking safety is an important research topic and a number of studies have been published to predict and reduce crash risk associated with trucks (Cantor et al., 2010; Chen et al., 2015; Dong et al., 2017).

Traditional crash prediction studies collect retrospective police reports of crashes in a given road section for a specified time period, match these crash cases with non-crash controls (typically 1 to 4 matching), and then build statistical models (such as logistic regression and neural networks) to study the risk factors associated with higher risk of crashes and predict real crashs (Blower et al., 2010; Meuleners et al., 2017; Sharwood et al., 2013). This case-control study design is efficient and less time-consuming in the field of trucking safety since crashes are very rare. However, case-control studies, by nature, are limited in study design since a) it is impossible to estimate and compare the rate of crashes since the number of non-crashes is unknown, b) retrospective reports are often subject to recall and report bias: the drivers may not accurately recall the exact conditions at the time of the event, c) the drivers may intentionally conceal some critical facts to escape from legal punishment (Dingus et al., 2011; Stern et al., 2019).

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

Naturalistic driving studies (NDSs) have been emerging in the past decade thanks to the advancement of technology. An NDS continuously collects driving data (including latitude, longitude, and speed) under real-world conditions using on-board unobtrusive equipment (Guo, 2019). In contrast to retrospective reports, an NDS resembles a cohort study: a pre-determined set of drivers are prospectively followed for a certain amount of time. Therefore, NDS has several advantages. First, NDS collects both crashes and non-crashes, so it is more useful in comparing the rates of events. Second, since vehicle crashes are extremely rare, it may take a huge amount of driving time to have sufficient sample of crashes. Instead, NDS focus safety-critical events (SCEs), which is defined as events that avoid crashes by last-second evasive maneuver (Dingus et al., 2011). SCEs can be 1000 times as high as real crashes and are argued to be good surrogates of crashes (Dingus et al., 2011; Guo et al., 2010; Johnsson et al., 2018; Mahmud et al., 2017). Third, NDS data are collected using programmed instruments or sensors, so they are less likely to be subject to human error, recall bias, or misinformation. Lastly, NDS collects data every a few seconds to minutes, and this large-scale high-resolution data provide a promising opportunity to quantifying driving risk (Guo, 2019).

However, many issues arise given the characteristics of NDSs. First, the sheer volume of NDS data creates a challenge to data management and aggregation (Mannering and Bhat, 2014). For example, a NDS data set can have billions rows of real-time speeds and locations, and it is important to have scalable and high-performance tools to aggregate these data into units that fit into the framework of statistical modeling. Second, routinely collected NDS data only have vehicle driving data. Crucial environmental variables such as weather and traffic need to be accessed from other data sources and merged back to the driving data. Third, even with these data sources, management, and aggregation issues solved, there is a lack of concensus on choosing the statistical models that are both sufficiently complex to account for the characteristics of NDS and computationally feasible to fit the large-scale data. With increasing companies collecting NDS data on a regular basis, a scalable and generalizable analyzing framework can serve as a pattern for researchers to better understand NDS data and gain insights into trucking and transportation safety.

This paper aims to propose a generalizable analytic framework (data collecting, aggregating, fusing, and statistical modeling) that accounts for the features of NDS data. To achive this aim, we have answers the following questions:

(A) How can we aggregate the high-resolutional NDS data into statistically analyzable units?

(B) Where are the third-party data sources available to improve statistical models?

(C) What are the risk factors associated with risky driving behavior among the sample truck drivers?

The remainder of this paper is organized as follows. Section 2 provides a brief literature review on previously published studies that use NDS datasets. Then, Section 3 presents our NDS data and other third-party data sources. Section 4 demonstrates how we aggregate the ping data into shifts, trips, and 30-minute intervals, and merge different sources of data. Statistical models, results and discussion, conclusion and implication.

2. Literature review

Although NDS data only emerge in the recent decade and are relatively new, there are an increasing number of data analytic studies published using this data. In this section, instead of exhaustively reviewing all published papers, we try to introduce a few relevant papers published in the recent years. The the data, methods, and results of these papers are briefly outlined and compared. Then we identify and summarize the research gaps, which are also the contribution and innovation that this study could make to the existing literature.

Using naturalistic driving data of 42 drivers in Shanghai, China, Zhu et al. (2018) compared five car-following models (Gazis-Herman-Rothery, Gipps, intelligent driver, full velocity difference, and Wiedemann). They found that the intelligent driver model had the best performance (lowest standard errors in calibration and validation) for the sample drivers.

Ghasemzadeh and Ahmed (2018) investigated the relationship between weather conditions and driver lane-keeping performance using 141 drivers from the second Strategic Highway Research Program (SHRP2). They used logistic regression and multivariate adaptive regression splines (MARS) to understand weather-related adverse conditions associated with lane-keeping behavior. Traffic conditions, driver's age and experience, posted speed limits were found to be significantly associated with driver lane-keeping ability.

Chen et al. (2016)

Soccolich et al. (2013)

Mollicone et al. (2019)

McCauley et al. (2013)

Sparrow et al. (2016)

Wu and Jovanis (2013)

Kamla et al. (2019)

Ahmed et al. (2018)

(A) A lack of studies specifically target truck drivers,

(B) The number of the drivers are small,

(C) No detailed description on data cleaning, aggregation, and fusion,

(D) Although hierarchical models are used, few studies actually used driver-centric hierarchical models.

3. Data sources

The data were collected by a leading freight shipping trucking company (we will name it as Company A for confidentiality reasons) in the United States. From April 2015 to March 2016, Company A equipped all their trucks with in-vehicle data acquisition systems (DAGs) that collect real-time *ping* and *SCEs* data. Details of these two data sources will be introduced in the following subsection. For demonstration purposes, we sampled 497 regional truck drivers who move freights in a region and surrounding states in this study. Apart from these vehicle driving data, demographic variables including age, gender, and race were also provided to the research team. The drivers were anonymized to ensure confidentiality, while a unique identification number was provided for each driver to link the three data sources. The study protocol was reviewed and approved by the Institutional Review Board of Saint Louis University.

3.1. Ping and SCEs data

The DAGs ping irregularly (typically every a couple of seconds to minutes) as the truck goes on road. Each ping collects several key variables, including the date and time (year, month, day, hour, minute, and second), latitude and longitude (specific to five decimal places), driver identification number (ID), and speed at that second. In total, 13,187,289 rows of ping data were generated by the 497 truck drivers.

Apart from ping data, Company A also collected real-time SCEs data for all their trucks. In contrast to irregularly collected ping data, SCEs were recorded whenever pre-determined kinematic thresholds were triggered. There were 9,032 critical events occurred to these 497 truck drivers during the study period. Four types of critical events were recorded in this critical events data, including 3,944 headway, 3,588 hard brakes, 869 collision mitigation, 631 rolling stability.

3.2. Weather

Apart from driver's characteristics and driving condition, weather also poses a threat on truck crashes and injuries (Naik et al., 2016; Uddin and Huynh, 2017; Zhu and Srinivasan, 2011). We obtained historic weather data from the DarkSky Application Programming Interface (API), which allows us to query historic real-time and hour-by-hour nationwide historic weather conditions according to latitude, longitude, date, and time (The Dark Sky Company, LLC, 2019). The variables included visibility, precipitation probability[1], precipitation intensity, temperature, wind, and others.

Traffic and road geometry can be collected from Google map API and OpenStreetmap API. However, querying historic traffic data for all our sample pings from Google map will create costs higher than the budget of the research team. The OpenStreetmap API is open-sourced and free platform that provides road geometry data (including speed limit and the number of lanes), but the missing rate ($> 50\%$) is too high to use for sample pings in this study. Therefore, we did not use traffic data or road geometry data in this study.

[1] Ideally, historic precipitation at a specific location and time should be yes or not. However, in reality, since the weather stations are distributed not densely enough to record the exact weather conditions in every latitude and longitude in the US, the DarkSky API uses their algorithms to infer the probability of precipitation in each location.

4. Data preparation

4.1. Shifts, trips, and 30-minute intervals

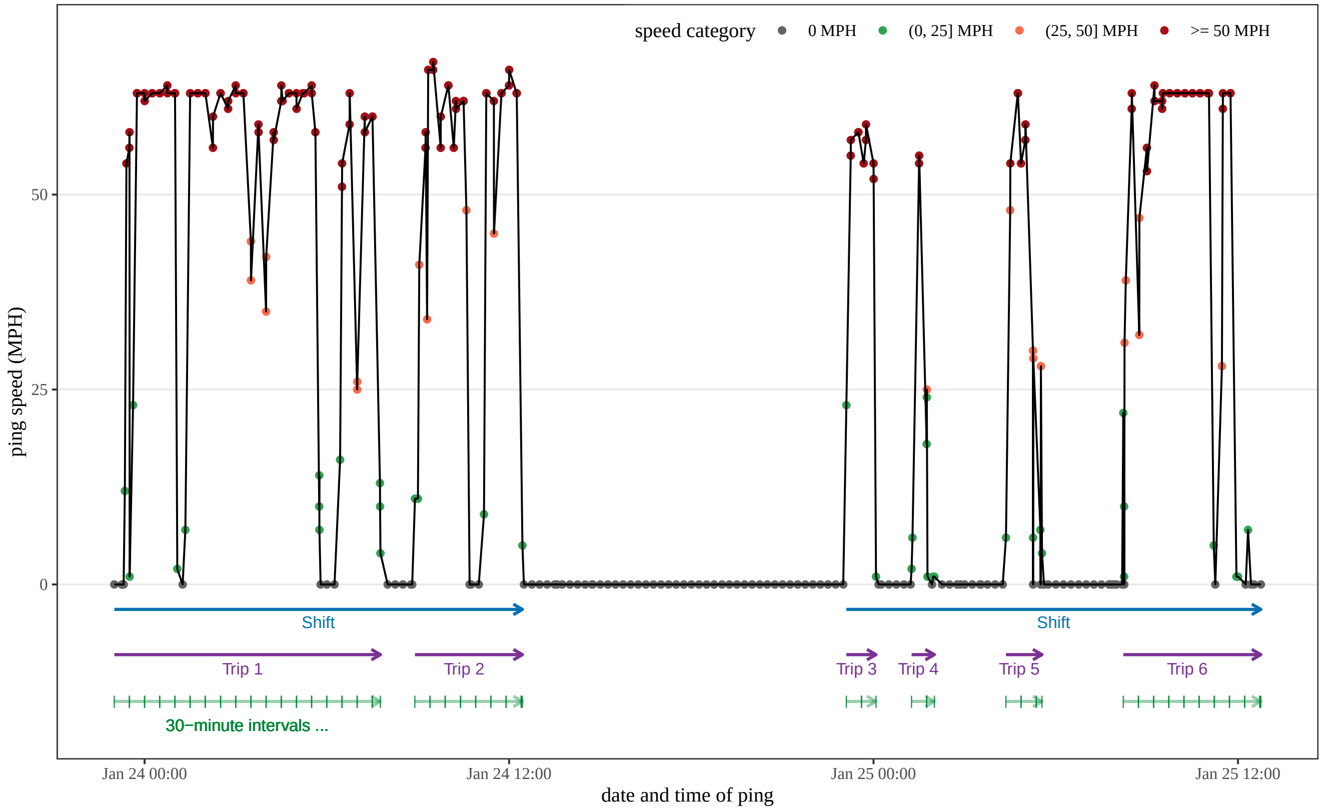


Figure 1: Data aggregation process from pings to shifts, trips, and 30-minute intervals.

To shrink the large size of over 10 million ping data, we rounded the GPS coordinates to the second decimal places, which are worth up to 1.1 kilometers, and we also round the time to the nearest hour. We then queried weather variables from the DarkSky API using the approximated latitudes, longitudes, date and hour. The weather variables used in this study include precipitation probability, precipitation intensity, and visibility.

For each of the truck drivers, if the ping data showed that the truck was not moving for more than 20 minutes, the ping data were separated into two different trips. These ping data were then aggregated into different trips. A **trip** is therefore defined as a continuous period of driving without stop. As Table demonstrates, each row is a trip. The average length of a trip in this study is 2.31 hours with the standard deviation of 1.8 hours.

After the ping data were aggregated into trips, these trips data were then further divided into different shifts according to an eight-hour rest time for each driver. A **shift** is defined as a long period of driving with potentially less than 8 hours' stops. The Shift_ID column in shows different shifts, separated by an eight-hour threshold. The average length of a shift in this study is 8.42 hours with the standard deviation of 2.45 hours.

4.2. Cumulative driving time as a measure of fatigue

Fatigue has been reported to be the most important predictor to truck crashes, considering that truck drivers are exposed to long routes and lone working environment Stern et al. (2019).

Driver's fatigue is difficult to measure in real life. In this study, we attempt to use cumulative driving time in a shift as a proxy measure of the fatigue of truck drivers.

4.3. Data fusion

5. Methodology

5.1. Statistical models

Traditional statistical models assume that observations are independent from each other given their predictor variables. However, natural data are almost never independent given the predictor variables. In the example of truck driver's safety events, if we assume the external traffic, weather and driver's socioeconomic status are fixed, truck drivers may exhibit similar driving patterns in multiple trips, and then drivers hired by the same company may share similar culture and safety atmospheres. Therefore, traffic accidents are naturally nested within drivers and drivers are nested within companies. Traditional statistical models that assume independence between observations are not appropriate in this case since objects tend to be similar within a group. Hierarchical models, also known as multilevel model, random-effects model or mixed model, have been developed to allow for the nested nature of data. Instead of assuming independence given predictor variables, hierarchical models assume conditional independence. Hierarchical models are advocated to be the default method since they can produce more precise prediction and more robust results than traditional models.

Random-effects models (Han et al., 2018; Pantangi et al., 2019).

Here we model the probability of a critical event occurred using two hierarchical models: logistic and negative binomial (NB) regression models. In the hierarchical logistic regression model, we categorized the number of safety events during the i-th 30-minute interval into a binary variable Y_i with the value of either 0 or 1, where 0 indicated that no critical event occurred during that trip while 1 indicated that at least 1 critical event occurred during the trip. The hierarchical logistic regression model is parameterized as:

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(p_i) \\ \log \frac{p_i}{1-p_i} &= \beta_{0,d(i)} + \beta_{1,d(i)} \cdot \text{CT}_i + \beta_2 x_2 + \cdots + \beta_k x_k \\ \beta_{0,d(i)} &\sim N(\mu_0, \sigma_0^2) \\ \beta_{1,d(i)} &\sim N(\mu_1, \sigma_1^2). \end{aligned} \tag{1}$$

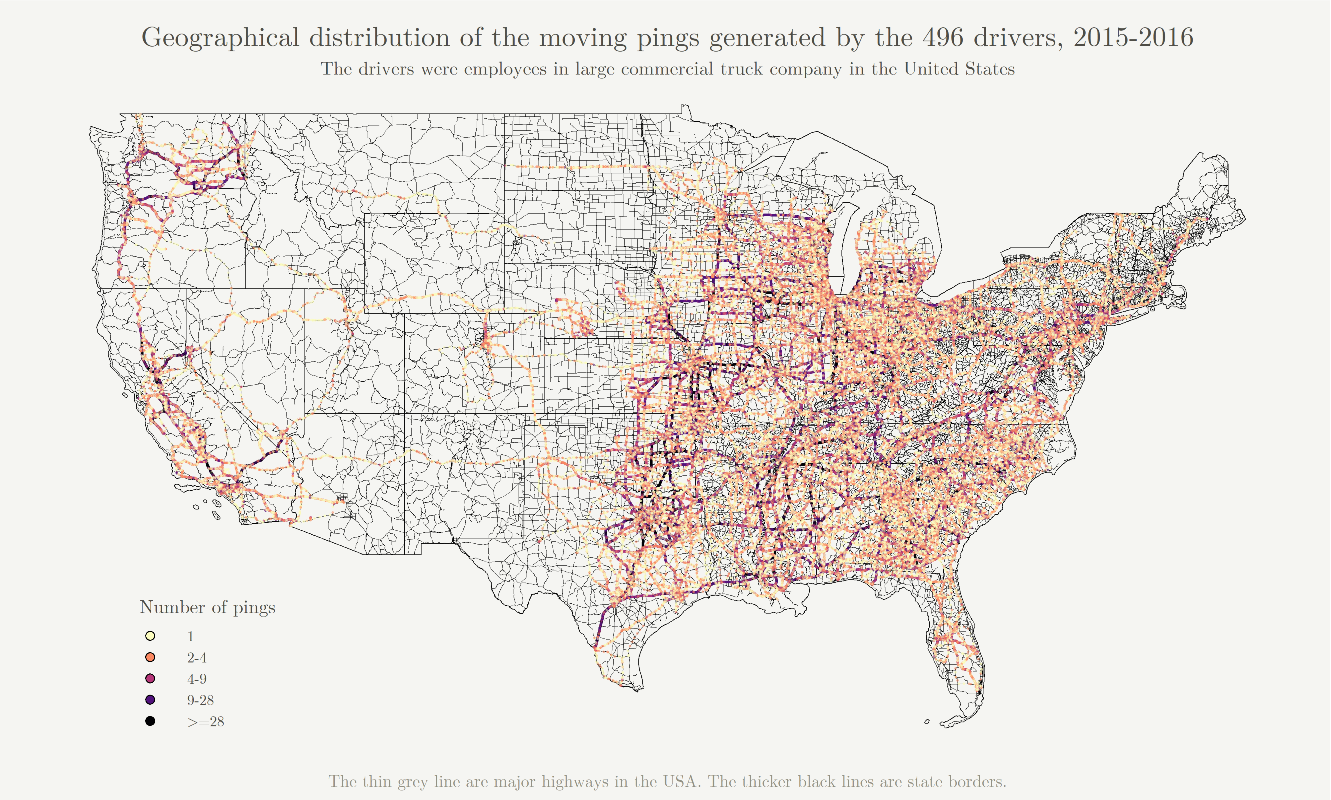
Here $d(i)$ is the driver for interval i, $\beta_{0,d(i)}$ is the random intercept for driver $d(i)$; $\beta_{1,d(i)}$ is the random slope for the cumulative driving time (CTi) in the shift (the sum of driving time for all previous intervals within that shift) for driver $d(i)$. These random intercepts and random slopes are assumed to have a hyper-distribution with hyperparameters $\mu_0, \sigma_0, \mu_1, \sigma_1$. $x_2, \ldots, x_k$ are other fixed-effect variables including driver demographics (age, gender, and race), weather (visibility, precipitation intensity and probability), interval specific variables (mean and standard deviation (s.d.) of speed), and $\beta_2, \ldots, \beta_k$ are the associated parameters.

Although logistic regression is more robust to outliers of the outcome variable in each 30-interval, it does not fully use the information in the outcome variable since only a binary variable is used. Here we present a hierarchical NB model, with the number of SCEs $Y_i^\star$ within the i-th interval as the outcome variable. The hierarchical NB regression model is parameterized as:

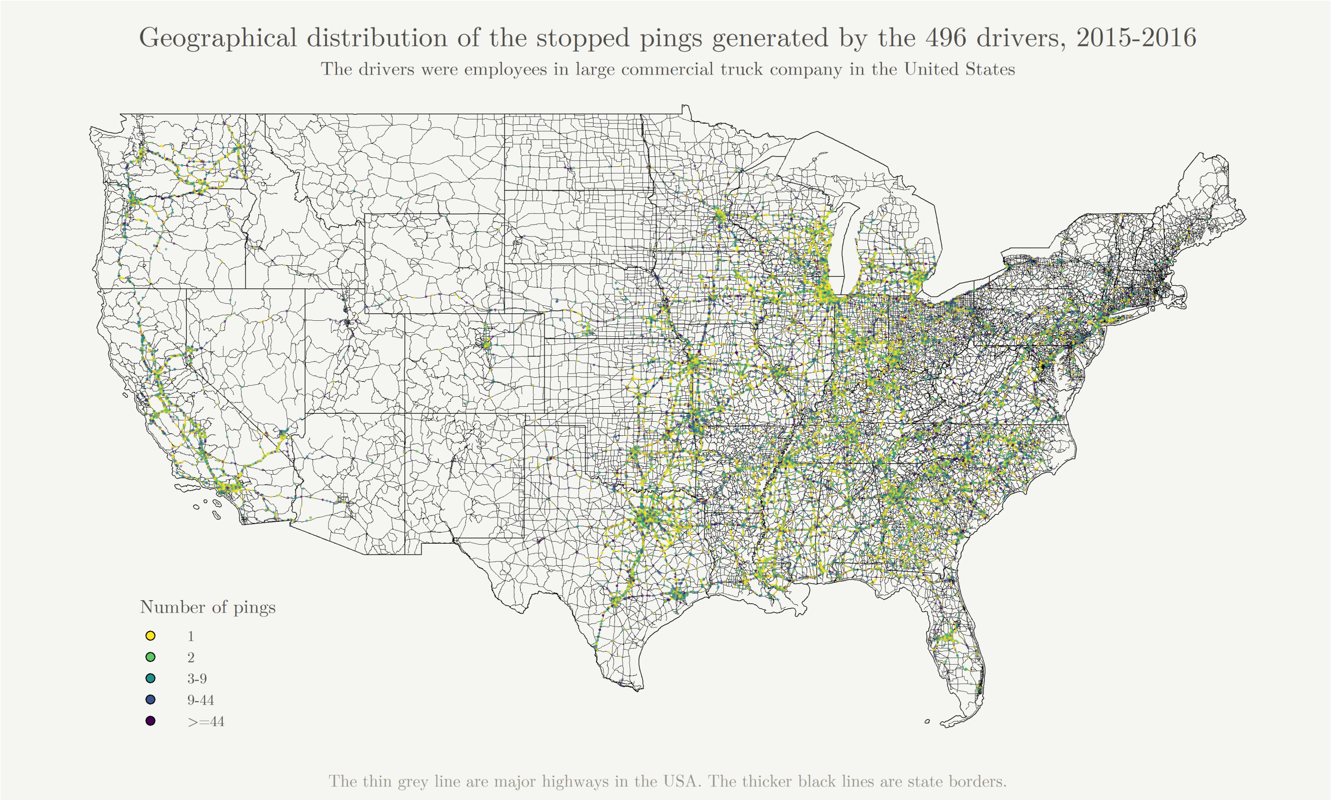
$$\begin{aligned} Y_i^\star &\sim \mathrm{NB}(T_i \times \mu_i,\ \mu_i + \frac{\mu_i^2}{\theta}) \\ \log \mu_i &= \beta^\star_{0,d(i)} + \beta^\star_{1,d(i)} \cdot \mathrm{CT}_i + \beta^\star_2 x_2 + \cdots + \beta^\star_k x_k \\ \beta^\star_{0,d(i)} &\sim N(\mu^\star_0, \sigma^{\star 2}_0) \\ \beta^\star_{1,d(i)} &\sim N(\mu^\star_1, \sigma^{\star 2}_1). \end{aligned} \tag{2}$$

Here T_i is the length of the i-th interval, μ_i is the expected number of SCEs per hour, θ is a fixed over-dispersion parameter. Other parameters are similar and explained in the previous hierarchical logistic regression model, and we put a $\star$ on the parameter to note the difference between the parameters of the two models.

The hierarchical logistic and NB models were estimated using the `lme4` package in R 3.6.2.



(a) Active pings



(b) Inactive pings

Figure 2: Geographical point patterns of moving and stopped pings generated by the 497 sample drivers.

6. Results and discussion

6.1. Sample description

6.2. Statistical models

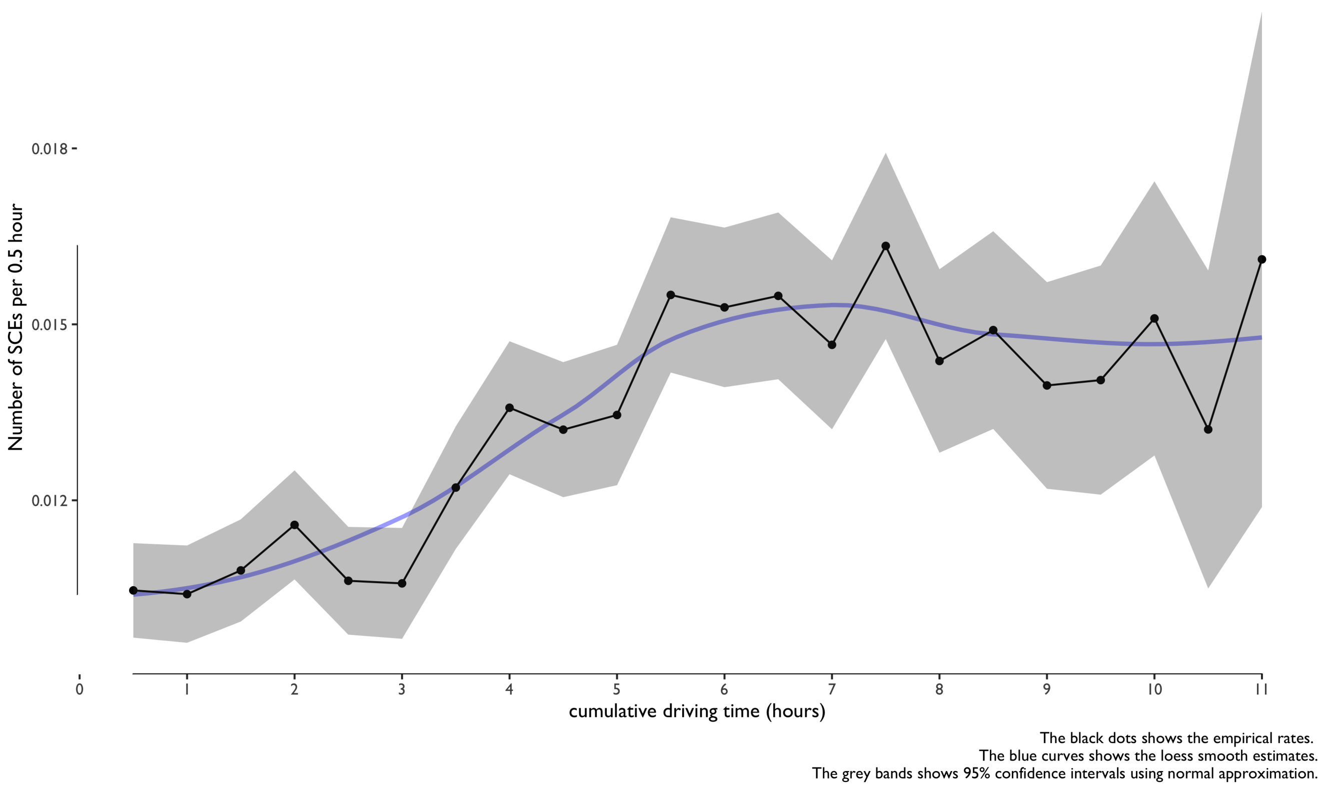


Figure 3: The rate of safety critical events and cumulative driving time

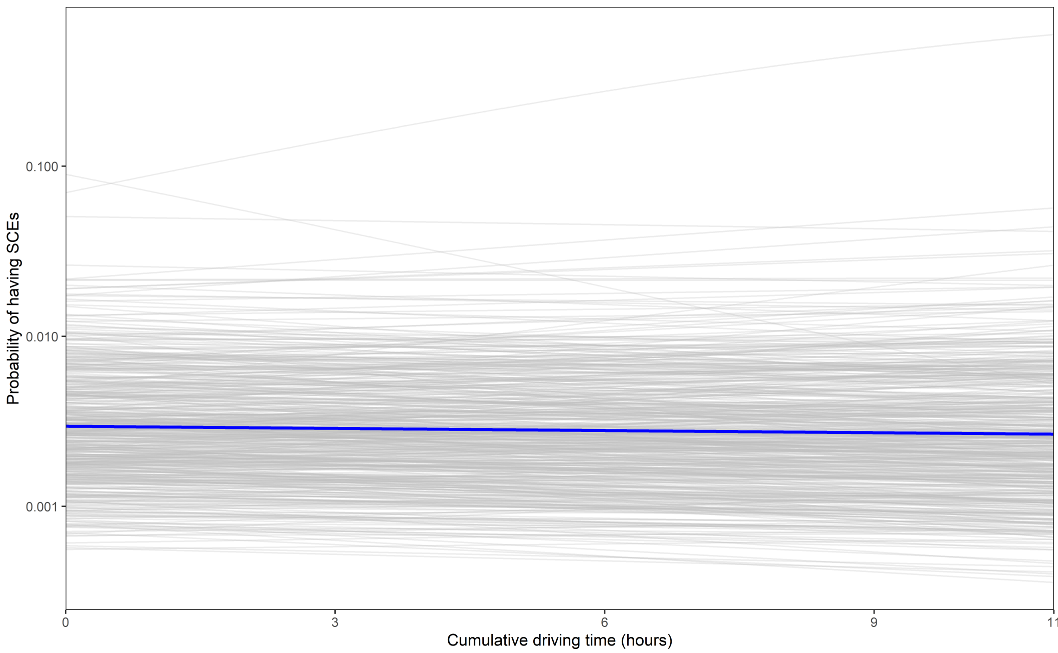
yardstick

7. Conclusions and implications

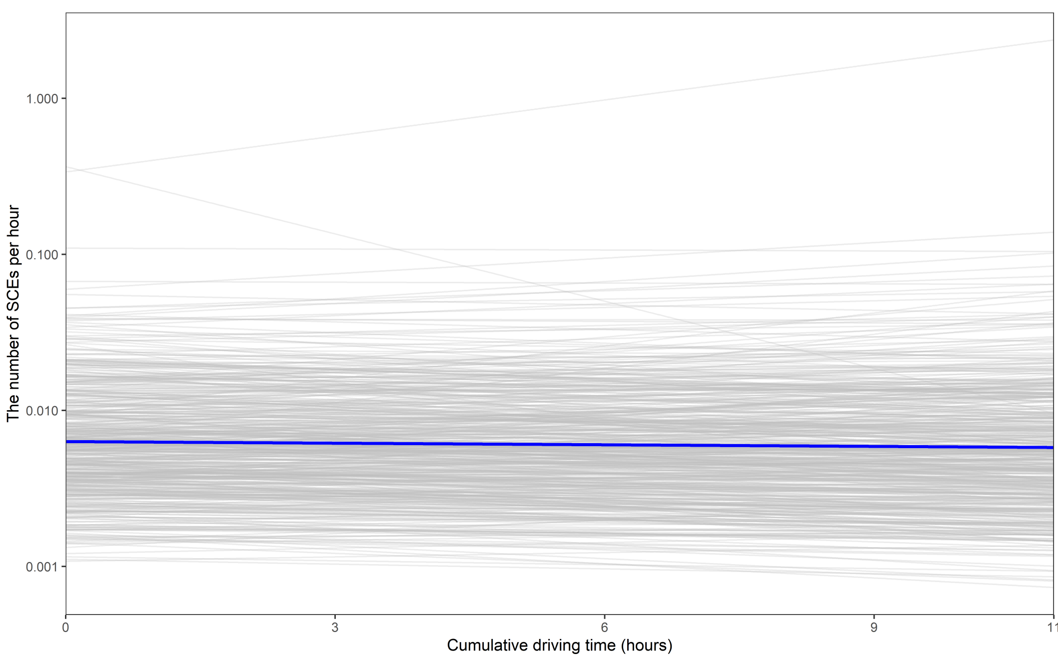
Although this study is based on NDS data generated from large commercial truck drivers, the driver-centric data collection, aggregation, fusing, and statistical modeling framework is generalizable to other types of drivers since the original ping and SCEs data are similar among different types of drivers.

Acknowledgement

This work was supported in part by the National Science Foundation (CMMI-1635927 and CMMI-1634992), the Ohio Supercomputer Center (PMIU0138 and PMIU0162), the American Society of Safety Professionals (ASSP) Foundation, the University of Cincinnati Education and Research Center Pilot Research Project Training Program,



(a) Hierarchical Logistics model



(b) Hierarchical negative binomial model

Figure 4: Simulated relationship between cumulative driving time and probability (logistics model)/rate (negative binomial model) of SCEs the 497 sample drivers.

Table 1: Standard and hierarchical logistic and NB models

	logistic	NB	hierarchical logistic	hierarchical NB
	(1)	(2)	(3)	(4)
Intercept	-4.979^{***}	-7.333^{***}	-5.819^{***}	-8.466^{***}
	(0.105)	(0.097)	(0.235)	(0.237)
Cumulative driving	-0.005	-0.004	-0.010	-0.008
	(0.004)	(0.004)	(0.006)	(0.007)
Mean speed	-0.0002	-0.0003	0.003^{***}	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Speed s.d.	0.020^{***}	0.017^{***}	0.023^{***}	0.020^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Age	-0.010^{***}	-0.016^{***}	-0.006	-0.007
	(0.001)	(0.001)	(0.004)	(0.004)
Race: black	-0.055^{**}	-0.124^{***}	0.091	0.093
	(0.025)	(0.026)	(0.105)	(0.108)
Race: other	0.238^{***}	0.145^{***}	0.369^{**}	0.347^{*}
	(0.042)	(0.046)	(0.179)	(0.186)
Gender: male	0.288^{***}	0.348^{***}		
	(0.050)	(0.053)		
Gender: female	0.064	0.061		
	(0.341)	(0.380)		
Precipitation intensity	0.519	0.418	0.997	0.961
	(0.663)	(0.704)	(0.670)	(0.662)
Precipitation probability	-0.175^{**}	-0.164^{**}	-0.024	0.059
	(0.072)	(0.075)	(0.074)	(0.073)
Wind speed	-0.011^{***}	-0.013^{***}	-0.023^{***}	-0.024^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
Visibility	-0.029^{***}	-0.043^{***}	0.011^{**}	0.010^{*}
	(0.005)	(0.005)	(0.006)	(0.006)
Interval time	0.015^{***}		0.017^{***}	
	(0.002)		(0.002)	
Observations	1,019,482	1,019,482	1,019,482	1,019,482
Log Likelihood	-46,303.850	-49,627.630	-43,042.570	-45,961.190
θ		0.036^{***} (0.001)		
Akaike Inf. Crit.	92,635.690	99,281.260	86,115.150	91,952.390
Bayesian Inf. Crit.			86,292.670	92,129.910

Note: *p<0.1; **p<0.05; ***p<0.01

and the Transportation Informatics Tier I University Transportation Center (TransInfo). We also thank the DarkSky company for providing us five million free calls to their historic weather API.

References

Ahmed, M.M., Franke, R., Ksaibati, K., Shinstine, D.S., 2018. Effects of truck traffic on crash injury severity on rural highways in wyoming using bayesian binary logit models. Accident Analysis & Prevention 117, 106–113.

Blower, D., Green, P.E., Matteson, A., 2010. Condition of trucks and truck crash involvement: Evidence from the large truck crash causation study. Transportation Research Record 2194, 21–28.

Cantor, D.E., Corsi, T.M., Grimm, C.M., Özpolat, K., 2010. A driver focused truck crash prediction model. Transportation Research Part E: Logistics and Transportation Review 46, 683–692.

Chen, C., Zhang, G., Tian, Z., Bogus, S.M., Yang, Y., 2015. Hierarchical bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accident Analysis & Prevention 85, 186–198.

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

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

Dong, C., Dong, Q., Huang, B., Hu, W., Nambisan, S.S., 2017. Estimating factors contributing to frequency and severity of large truck–involved crashes. Journal of Transportation Engineering, Part A: Systems 143, 04017032.

Ghasemzadeh, A., Ahmed, M.M., 2018. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: Non-parametric mars and parametric logistic regression modeling approaches. Transportation research part C: emerging technologies 90, 379–392.

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

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.

Han, C., Huang, H., Lee, J., Wang, J., 2018. Investigating varying effect of road-level factors on crash frequency across regions: A bayesian hierarchical random parameter modeling approach. Analytic methods in accident research 20, 81–91.

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.

Johnsson, C., Laureshyn, A., De Ceunynck, T., 2018. In search of surrogate safety indicators for vulnerable road users: A review of surrogate safety indicators. Transport reviews 38, 765–785.

Kamla, J., Parry, T., Dawson, A., 2019. Analysing truck harsh braking incidents to study roundabout accident risk. Accident Analysis & Prevention 122, 365–377.

Mahmud, S.S., Ferreira, L., Hoque, M.S., Tavassoli, A., 2017. Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. IATSS research 41, 153–163.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research 1, 1–22.

McCauley, P., Kalachev, L.V., Mollicone, D.J., Banks, S., Dinges, D.F., Van Dongen, H.P., 2013. Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. Sleep 36, 1987–1997.

Meuleners, L., Fraser, M.L., Govorko, M.H., Stevenson, M.R., 2017. Determinants of the occupational environment and heavy vehicle crashes in western australia: A case–control study. Accident Analysis & Prevention 99, 452–458.

Mollicone, D., Kan, K., Mott, C., Bartels, R., Bruneau, S., Wollen, M. van, Sparrow, A.R., Van Dongen, H.P., 2019. Predicting performance and safety based on driver fatigue. Accident Analysis & Prevention 126, 142–145.

Naik, B., Tung, L.-W., Zhao, S., Khattak, A.J., 2016. Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research 58, 57–65.

National Highway Traffic Safety Administration, 2017. A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System.

Pantangi, S.S., Fountas, G., Sarwar, M.T., Anastasopoulos, P.C., Blatt, A., Majka, K., Pierowicz, J., Mohan, S.B., 2019. A preliminary investigation of the effectiveness of high visibility enforcement programs using naturalistic driving study data: A grouped random parameters approach. Analytic Methods in Accident Research 21, 1–12.

Sharwood, L.N., Elkington, J., Meuleners, L., Ivers, R., Boufous, S., Stevenson, M., 2013. Use of caffeinated substances and risk of crashes in long distance drivers of commercial vehicles: Case-control study. BMJ 346, f1140.

Soccolich, S.A., Blanco, M., Hanowski, R.J., Olson, R.L., Morgan, J.F., Guo, F., Wu, S.-C., 2013. An analysis of driving and working hour on commercial motor vehicle driver safety using naturalistic data collection. Accident Analysis & Prevention 58, 249–258.

Sparrow, A.R., Mollicone, D.J., Kan, K., Bartels, R., Satterfield, B.C., Riedy, S.M., Unice, A., Van Dongen, H.P., 2016. Naturalistic field study of the restart break in us commercial motor vehicle drivers: Truck driving, sleep, and fatigue. Accident Analysis & Prevention 93, 55–64.

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., others, 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 Dark Sky Company, LLC, 2019. Dark Sky API — Overview.

Uddin, M., Huynh, N., 2017. Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accident Analysis & Prevention 108, 44–55.

WHO, 2018. The top 10 causes of death.

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

Zhu, M., Wang, X., Tarko, A., others, 2018. Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study. Transportation research part C: emerging technologies 93, 425–445.

Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Analysis & Prevention 43, 49–57.