Modeling Truck Safety Critical Events: Efficient Bayesian Hierarchical Statistical and Reliability Models

(Dissertation defense 2020)

Miao Cai

Department of Epidemiology and Biostatistics,

Saint Louis University

July 20, 2020

SAINT LOUIS UNIVERSITY

COLLEGE FOR PUBLIC HEALTH
AND SOCIAL JUSTICE

Dissertation Committee

Steven E. Rigdon
Saint Louis University

Committee chair

Hong Xian
Saint Louis University

Committee member

Fadel Megahed
Miami University

Committee member

1 Introduction

Commercial Truck Drivers

- They "form the lifeblood of [the U.S.] economy"[1]:
  - Annual revenues of > 700 billion USD[2]
  - 10.8 billion tons of freight[2].
- A labor intensive industry:
  - Long routes
  - Under on-time demands
  - Late-night or early-morning shifts → sleep deprivation/disorder.
- A major public health threat:
  - Large physical dimensions and potential hazardous cargoes
  - Fatal crashes: 4,564 people killed in 4,079 crashes[3]
  - 78% of deaths occurred in the other vehicles[4]
  - ≈91,000 USD per crash[5].

Traditional Trucking Safety Studies

Traditional trucking safety studies utilize road segments as the unit of analysis[6]. They collect the crashes on road segments and match them with non-crashes. The *occurrence* or the *number of crashes* are modeled as the outcomes.

This case-control study design has several limitations:[6;7;8]:

- Non-crashes are not fully studied
- Difficulty in selecting control groups
- An extremely small number of observed crashes
- Undercount of less severe crashes.

Naturalistic Driving Studies (NDSs)

NDSs use unobtrusive devices, sensors, and cameras to proactively collect frequent naturalistic driving behavior and performance data under real-world driving conditions[7;9].

Comparatively, this prospective cohort study design has several strengths:[6]

- Events and non-events are all collected
- High-resolution data
- Less costly per observation

However, NDSs are relatively new. No clear and consistent framework for data aggregation, reduction, statistical modeling for this type of data.

Safety-Critical Events (SCEs)

Safety-critical events (SCEs) are special types of accident precursors that have all features of accidents, except that potentially catastrophic outcomes were avoided by last-second evasive maneuvers [10;11]

Examples of SCEs: [6]

- Hard brakes
- Headway
- Collision mitigation
- Rolling stability.

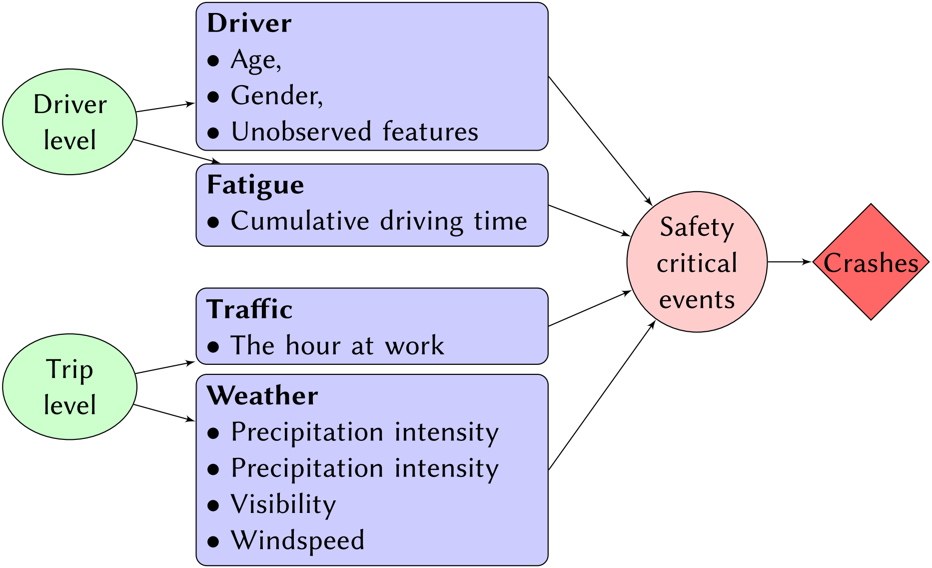
No studies have examined the association between crashes and surrogates using NDS data sets that specifically target commercial truck drivers.

Research Aims

Overarching goal: to construct a generalizable analysis framework for trucking NDS data and understand driving behavior.

1. Aim 1: To examine the association between truck crashes and SCEs among commercial truck drivers.
   - Bayesian negative binomial regression models
2. Aim 2: To aggregate NDS data into analyzable units and build two scalable hierarchical models.
   - Hierarchical logistic regression models
   - Hierarchical negative binomial regression models
3. To study the pattern of SCEs within shifts using Bayesian recurrent event models.
   - A Bayesian hierarchical non-homogeneous Poisson process with power law process intensity function
   - A Bayesian hierarchical jump power law process to account for within-shift breaks.

Conceptual Framework



2 Aim 1

2 Aim 1

2.1 Methods

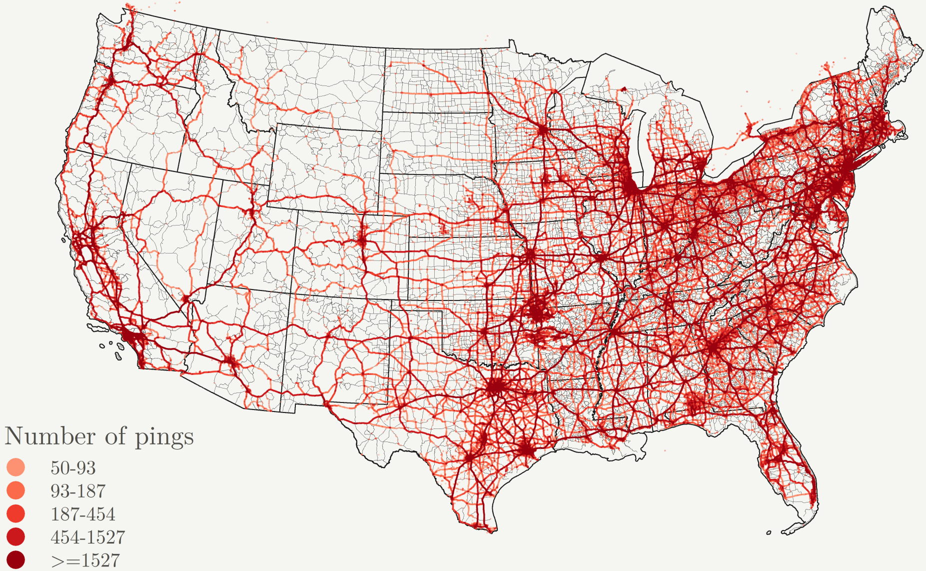
Data Sets

31,828 commercial truck drivers, covering 2,320,967,467 miles and 65,646,731 hours.

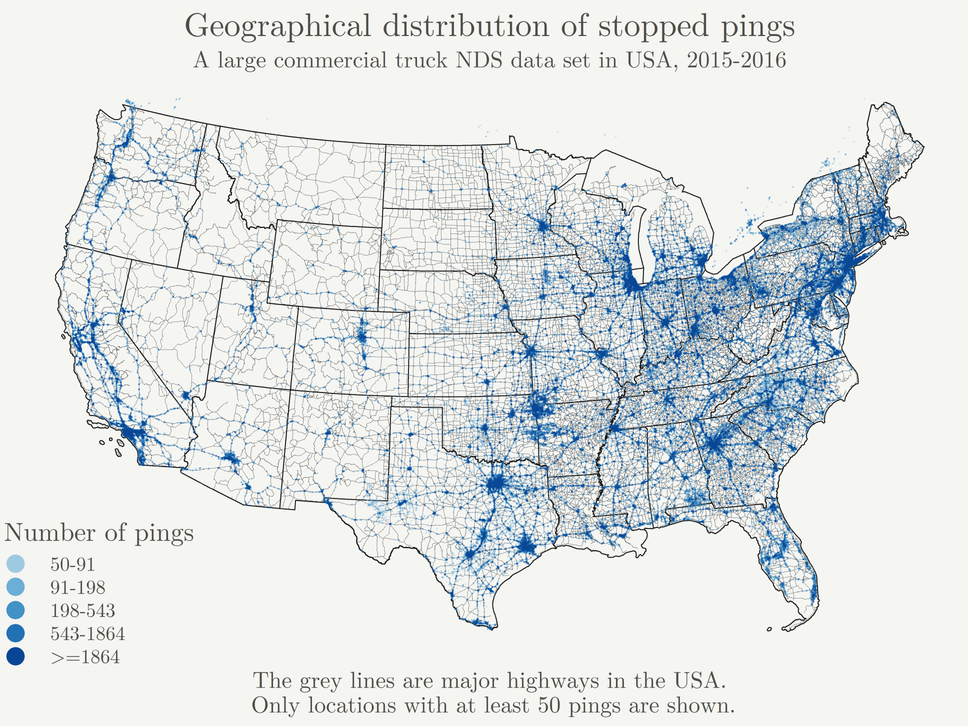
- Pings: 1,494,678,173 pings
  - Date, time, speed, and GPS (latitudes and longitudes)
- SCEs:
  - 170,421 headway,
  - 218,419 hard brakes,
  - 55,243 forward-collision mitigation,
  - 6,675 initiations of the rolling stability system.
- Crashes:
  - 34,884 crashes, 239 injuries, and 22 fatalities
- Driver characteristics:
  - Age, race, gender
  - Driver type: local, regional, over-the-road
  - Business units: dedicated, intermodal, final-mile.

Geographical distribution of active moving pings

A large commercial truck NDS data set in USA, 2015-2016

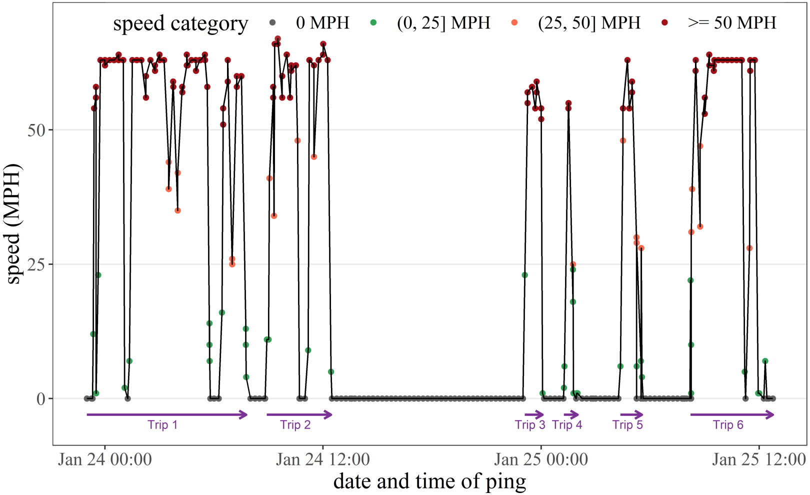


The grey lines are major highways in the USA.
Only locations with at least 50 pings are shown.


Geographical distribution of stopped pings
A large commercial truck NDS data set in USA, 2015-2016
Number of pings
50-91
91-198
198-543
543-1864
>=1864
The grey lines are major highways in the USA.
Only locations with at least 50 pings are shown.

Data Aggregation

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.[12]



Bayesian Negative Binomial (NB) models

$$\begin{aligned} Y_i &\sim \text{Negative Binomial}(T_i \times \mu_i,\ \phi) \\ \log \mu_i &= \alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_J x_{iJ} + \theta_1 z_{i1} + \cdots + \theta_K z_{iK}, \end{aligned} \tag{1}$$

- Outcomes Y_i: # of crashes, injuries, or fatalities per driver
- Offset T_i: total # of miles driven per driver
- Overdispersion parameter ϕ
- Predictors of interest $x_{i1}, x_{i2}, x_{i3}, x_{i4}$:
  - # of headway, hard brakes, collision mitigation, and rolling stability per 10,000 miles
- Covariates z_{ik}:
  - Age, gender, mean speed
  - Driver types, business units.

Four Markov chains per model, 4,000 iterations (including 1,000 warmup steps) per chain.[13]

2 Aim 1

2.2 Results

SCEs Associated with Crashes, Injuries, and Fatalities

Variables	Crashes: pooled	Crashes: four SCEs	Injuries: pooled	Injuries: four SCEs	Fatalities: pooled	Fatalities: four SCEs
All SCEs	1.084 (1.080, 1.088)		1.087 (1.048, 1.136)		0.973 (0.791, 1.149)	
Headway		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)
Fit statistics:						
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.

Consistent Results among Four Different SCEs

Variables	Pooled	Four SCEs	Headway	Hard brakes	Rolling stability	Collision mitigation
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)
Fit statistics:						
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.

Stratified by Business Units and Driver Types

Variables	Dedicated Local	Dedicated OTR	Dedicated Regional	Inter-modal Local	Inter-modal Regional	Final-mile OTR	Final-mile Regional
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)
Fit statistics:							
Sample size	6950	1797	7405	6429	3339	943	4963
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. OTR: over-the-road drivers.

Model Validation and Diagnostics

		Hard brake	Rolling stability	Collision mitigation	Age	Gender	Ping speed	Business unit
Hard brake	0.14							
Rolling stability	0.11	0.12						
Collision mitigation	0.19	0.13	0.17					
Age	−0.02	−0.04	−0.04	−0.02				
Gender	0.03	0.02	0.03	0.03	−0.01			
Ping speed	0.02	−0.14	−0.03	−0.06	0.02	0		
Business unit	−0.01	0.01	−0.01	0.02	0	−0.01	−0.07	
Driver type	0	−0.11	−0.03	−0.06	0.08	0.01	0.6	0.2

- Besides, all variance inflation factors are $\leq$1.3, indicating no evidence of multicollinearity issues.[14]

3 Aim 2

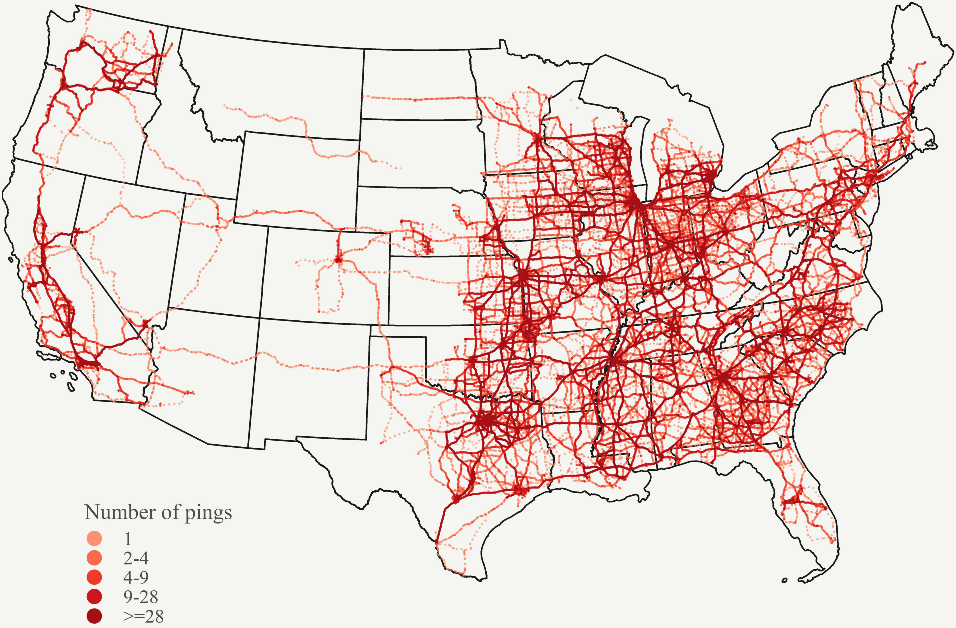
3 Aim 2

3.1 Methods

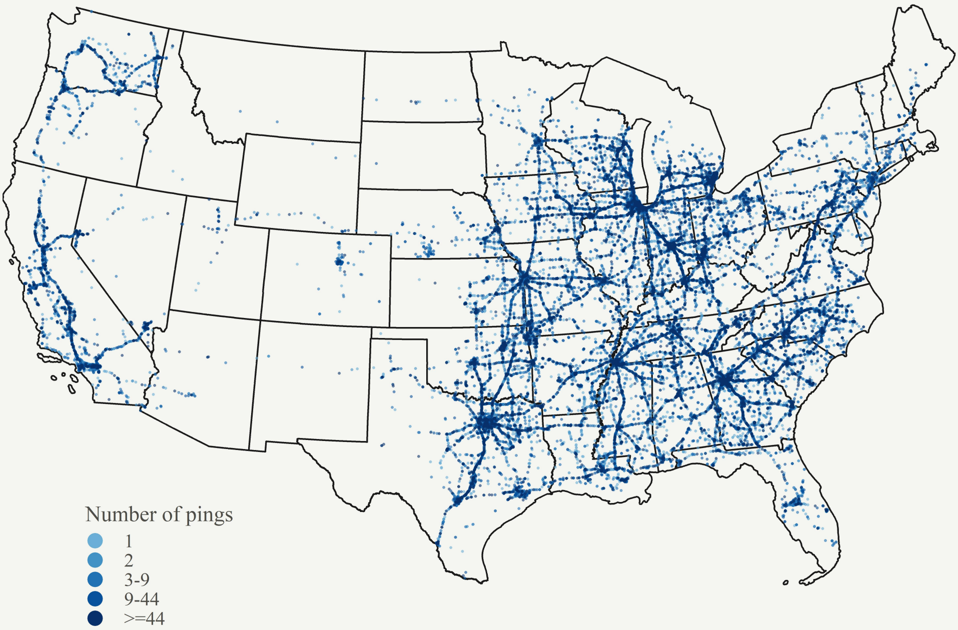
Data Sets

496 commercial truck drivers data set.

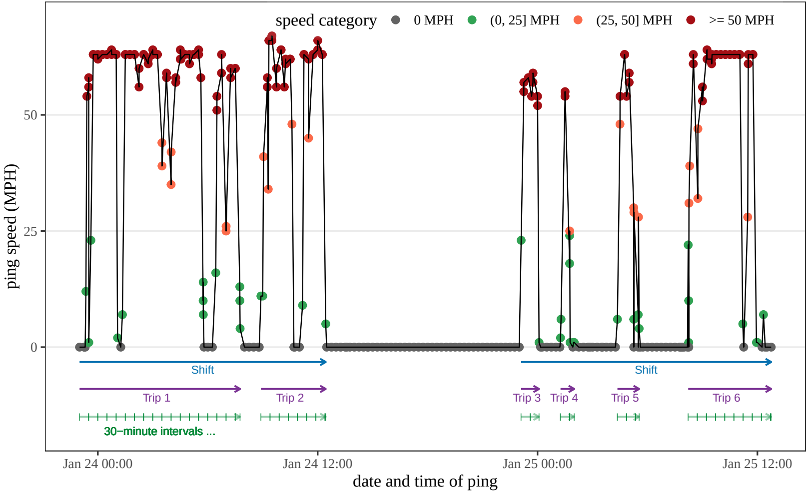
- Pings: 13,187,289 pings (≈13 million)
  - Date, time, speed, and GPS (latitudes and longitudes)
  - Covering 20,042,519 miles (≈20 million)
  - Spanning 465,641 hours.
- SCEs:
  - 3,941 headway,
  - 3,576 hard brakes,
  - 869 forward-collision mitigation,
  - initiations of the rolling stability system.



Number of pings
1
2-4
4-9
9-28
>=28


Number of pings
1
2
3-9
9-44
>=44

Data Aggregation



≈13M pings → 64,860 shifts → 180,408 trips → 1,018,475 intervals.

A Summary of Aggregated Data (496 drivers)

- 246 (49.6%) white, 206 (41.5%) black, and 44 (8.9%) other race.
- Female: 36 (7.3%). Average age 45.86 (12.01).

Category	Variables	Overall
	30-minute interval level variables (N=1,018,475)	
Weather	Precipitation intensity (mean (SD))	0.00 (0.02)
	Precipitation probability (mean (SD))	0.06 (0.21)
	Wind speed (mean (SD))	3.74 (3.20)
	Visibility (mean (SD))	8.74 (2.23)
Traffic proxy	Weekend = Yes	150,356 (14.8%)
	Holiday = Yes	15,348 (1.5%)
	Hour of the day	
	9 p.m. - 5 a.m.	123,460 (12.1%)
	6 a.m. - 10 a.m.	273,364 (26.8%)
	11 a.m. - 2 p.m.	294,058 (28.9%)
	3 p.m. - 8 p.m.	327,593 (32.2%)
Interval	Cumulative driving hours (mean (SD))	4.47 (2.76)
	Speed mean (mean (SD))	43.01 (19.77)
	Speed standard deviation (mean (SD))	11.22 (10.37)
	Interval time in minutes (mean (SD))	27.40 (6.70)

Hierarchical Logistic Regression

Y_i is a binary variable of if any SCEs occurred in the ith interval.

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

- $d(i)$: index of the driver for the ith interval
- $\beta_{0,d(i)}$: random intercept for driver $d(i)$
- CT_i: cumulative driving time in the shift
- $\beta_{1,d(i)}$: random slope of CT_i for driver $d(i)$
- $x_2, \ldots, x_k$: other fixed-effect variables
- $\mu_0, \sigma_0, \mu_1, \sigma_1$: hyperparameters.

Hierarchical Negative Binomial Regression

$Y_i^\star$ is the number of SCEs in the ith interval.

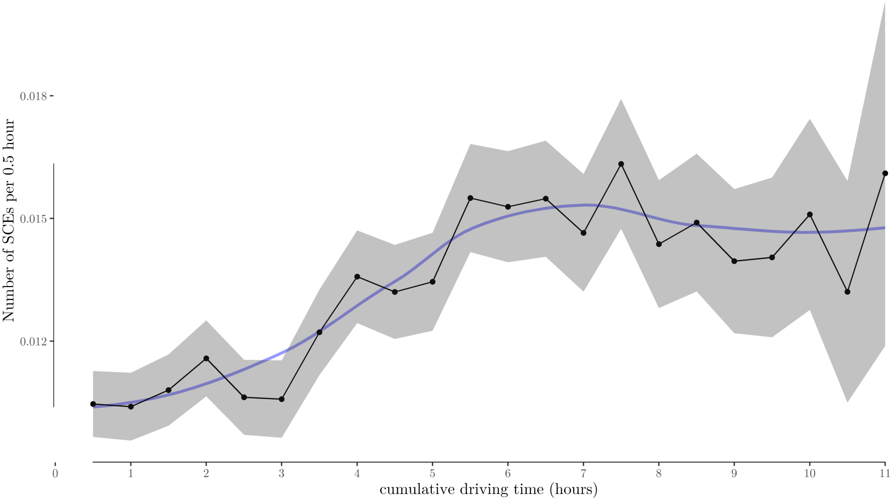
$$
\begin{aligned}
Y_i^\star &\sim \text{Negative Binomial}(T_i \times \mu_i,\ \mu_i + \frac{\mu_i^2}{\theta}) \\
\log \mu_i &= \beta_{0,d(i)}^\star + \beta_{1,d(i)}^\star \cdot \mathrm{CT}_i + \beta_2^\star x_2 + \cdots + \beta_k^\star x_k \\
\beta_{0,d(i)}^\star &\sim N(\mu_0^\star, \sigma_0^{\star 2}) \\
\beta_{1,d(i)}^\star &\sim N(\mu_1^\star, \sigma_1^{\star 2}),
\end{aligned}
\tag{3}
$$

- T_i: the length of the ith interval
- μ_i: the expected number of SCEs per interval
- θ: a fixed over-dispersion parameter.

3 Aim 2

3.2 Results

Cumulative Driving Time and Rate of SCEs

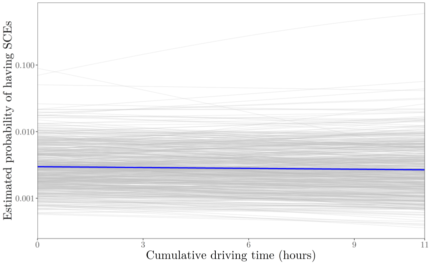


The black dots shows the empirical rates.
The blue curves shows the loess smooth estimates.
The grey bands shows 95% confidence intervals using normal approximation.

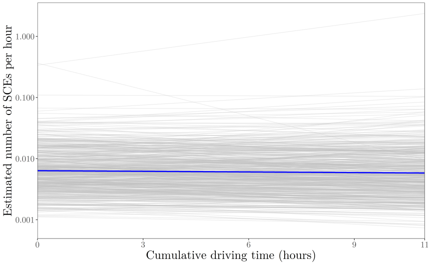
Logistic and NB models Modeling SCEs

	Logistic	NB	Hierarchical logistic	Hierarchical NB
	(1)	(2)	(3)	(4)
Intercept (μ_0)	-4.924***	-7.175***	-6.229***	-8.872***
	(0.099)	(0.090)	(0.240)	(0.242)
Cumulative driving hours (μ_1)	-0.006	-0.011**	-0.005	-0.005
	(0.005)	(0.005)	(0.007)	(0.007)
Mean speed	-0.00004	-0.0002	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)
Weekend	-0.144***	-0.187***	-0.119***	-0.117***
	(0.033)	(0.035)	(0.035)	(0.035)
Holiday	-0.315***	-0.339***	-0.351***	-0.345***
	(0.108)	(0.111)	(0.108)	(0.109)
Observations	1,019,482	1,019,482	1,019,482	1,019,482
θ		0.037***(0.001)		0.151
sd: Intercept (σ_0)			0.955	1.00
sd: cumulative driving (σ_1)			0.076	0.080
cor: μ_0 & μ_1			-0.192	-0.225
Note:				*p<0.1; **p<0.05; ***p<0.01

Simulated Random Intercepts & Slopes



Hierarchical Logistics model



Hierarchical negative binomial model

Simulated relationship between cumulative driving time and probability (logistics model)/rate (negative binomial model) of SCEs the 497 sample drivers. The y-axises are on the log 10 scale.

Hierarchical Logistic Regressions for Different SCEs Types

	Hard brake	Headway	Collision mitigation
	(1)	(2)	(3)
Intercept (μ_0)	-5.377***	-10.542***	-8.193***
	(0.271)	(0.410)	(0.408)
Cumulative driving hours (μ_1)	-0.0003	0.005	-0.005
	(0.009)	(0.011)	(0.017)
Mean speed	-0.018***	0.038***	-0.020***
	(0.001)	(0.002)	(0.002)
Speed s.d.	0.039***	0.028***	0.029***
	(0.002)	(0.002)	(0.003)
Weekend	0.104**	-0.388***	-0.081
	(0.049)	(0.057)	(0.103)
Holiday	-0.071	-0.621***	-0.462
	(0.152)	(0.171)	(0.356)
Observations	1,019,482	1,019,482	1,019,482
sd: Intercept (σ_0)	0.999	1.61	0.929
sd: cumulative driving (σ_1)	0.071	0.088	0.081
cor: μ_0 & μ_1	-0.285	-0.422	-0.276

Model Fit Statistics

Model	Log likelihood	AIC	BIC	$c-$statistic	Accuracy	Sensitivity	Specificity
Logistic	$-46{,}181$	92,397	92,610	0.602	0.285	0.845	0.281
NB	$-49{,}500$	99,035	99,248	0.585			
Hierarchical logistic	$-42{,}895$	85,832	86,080	0.770	0.593	0.795	0.591
Hierarchical NB	$-45{,}830$	91,701	91,950	0.700			

Model	Log likelihood	AIC	BIC	c-statistic	Accuracy	Sensitivity	Specificity
Hard brake	$-20{,}435$	40,911	41,160	0.800	0.654	0.829	0.654
Headway	$-20{,}491$	41,024	41,272	0.860	0.710	0.860	0.709
Collision mitigation	$-6{,}593$	13,228	13,476	0.830	0.614	0.872	0.614

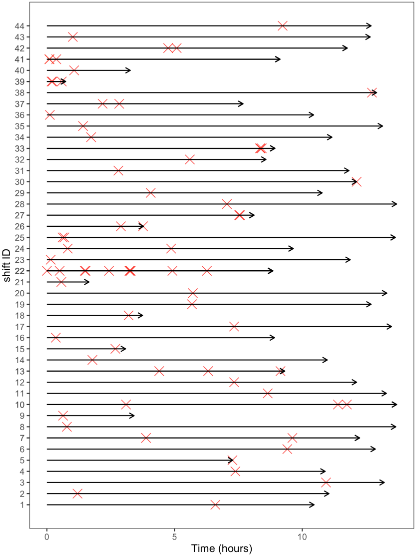
4 Aim 3

4 Aim 3

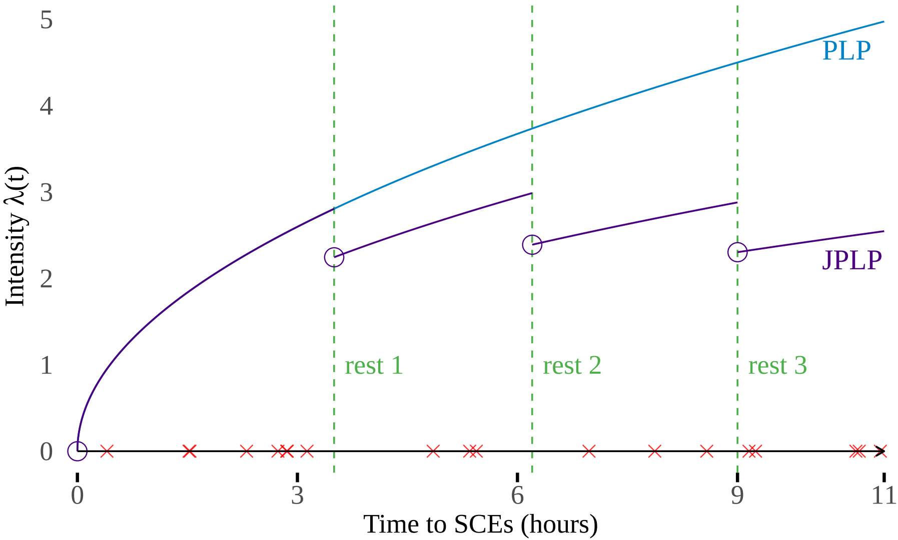
4.1 Methods

Motivation for Recurrent Events Models

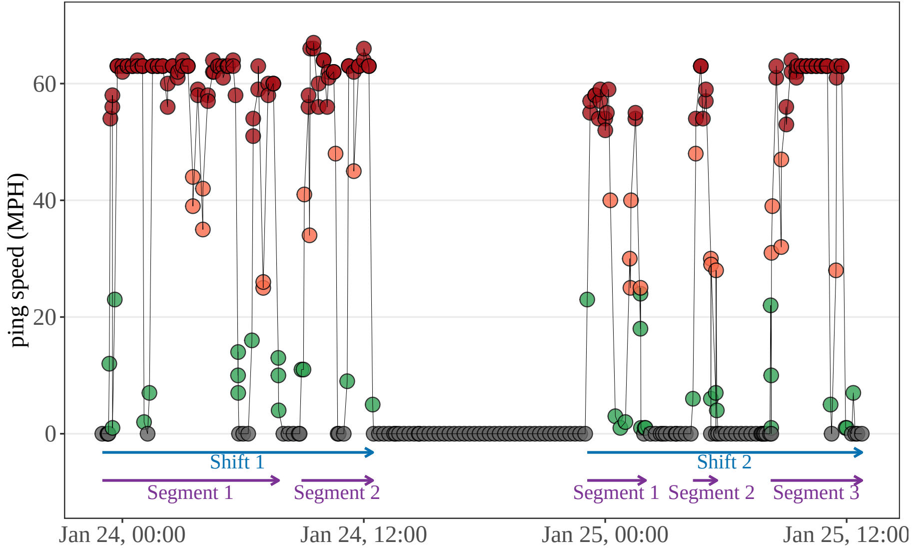
- A crash would end the shift
- SCEs:
    - Multiple SCEs can occur in one shift
    - SCEs do not interrupt the state of driving



Power Law Process (PLP) and Jump JPLP (JPLP)



Data Aggregation



Intensity Functions of PLP and JPLP

- Intensity function for a PLP

$$\lambda_{\text{PLP}}(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \tag{4}$$

where β is the shape parameter, θ is the rate parameter.

- Intensity function for a JPLP:

$$\begin{aligned}
&\lambda_{\text{JPLP}}(t|d, s, r, \beta, \gamma_{0d}, \gamma, \mathbf{X}_d, \mathbf{W}_s) = \\
&\quad\begin{cases}
\kappa^0 \lambda(t|\beta, \gamma_{0d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & 0 < t \leq a_{d,s,1}, \\
\kappa^1 \lambda(t|\beta, \gamma_{0d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & a_{d,s,1} < t \leq a_{d,s,2}, \\
\vdots & \vdots \\
\kappa^{R-1} \lambda(t|\beta, \gamma_{0d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & a_{d,s,R-1} < t \leq a_{d,s,R},
\end{cases} \\
&\quad = \kappa^{r-1} \lambda(t|d, s, r, \kappa, \beta, \gamma_{0d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), \quad a_{d,s,r-1} < t \leq a_{d,s,r},
\end{aligned} \tag{5}$$

where κ is an additional jump parameter.

Parameterization of Hierarchical PLP and JPLP

- Parameterization of a hierarchical PLP

$$\begin{aligned}
\left(t_{d,s,1}, t_{d,s,2}, \ldots, t_{d,s,n_{d,s}}\right) &\sim \text{PLP}(\beta, \theta_{d,s}, \tau_{d,s}) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \cdots + \gamma_k x_{d,s,k} \\
\gamma_{01}, \gamma_{02}, \ldots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2),
\end{aligned} \tag{6}$$

- Parameterization of a hierarchical JPLP

$$\begin{aligned}
\left(t_{d,s,1}, t_{d,s,2}, \ldots, t_{d,s,n_{d,s}}\right) &\sim \text{JPLP}(\beta, \theta_{d,s}, \tau_{d,s}, \kappa) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \cdots + \gamma_k x_{d,s,k} \\
\gamma_{01}, \gamma_{02}, \ldots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2).
\end{aligned} \tag{7}$$

- $t_{d,s,i}$: time to the i-th event for driver d in shift s,
- $\tau_{d,s} = a_{d,s,R_{d,s}}$: length of time of shift s (truncation time) for driver d,
- $n_{d,s} = \sum_{r=1}^{n_{d,s}}$: number of SCEs in shift s for driver d.

Priors and Hyperpriors of Hierarchical PLP and JPLP

Relatively non-informative priors for parameters for both models:

$$\begin{aligned} \beta &\sim \text{Gamma}(1,1) \\ \gamma_1, \gamma_2, \ldots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\ \mu_0 &\sim N(0, 5^2) \\ \sigma_0 &\sim \text{Gamma}(1,1). \end{aligned} \tag{8}$$

$$\begin{aligned} \beta &\sim \text{Gamma}(1,1) \\ \kappa &\sim \text{Uniform}(0,2) \\ \gamma_1, \gamma_2, \ldots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\ \mu_0 &\sim N(0, 5^2) \\ \sigma_0 &\sim \text{Gamma}(1,1). \end{aligned} \tag{9}$$

- $\kappa \sim \text{Uniform}(0, 2)$ to allow for potential jump up and down,
- Variance of normal priors are recommended by Stan.[13]

Likelihood Functions of Hierarchical PLP and JPLP

- Likelihood function of a hierarchical PLP

$$
\begin{aligned}
& L_{d,s}(\beta, \gamma_{0d}, \vec{\gamma} | \mathbf{X}_d, \mathbf{W}_s) \\
& = \begin{cases} \exp\Big(-(\tau_{d,s}/\theta_{d,s})^{\beta}\Big), & \text{if } n_{d,s} = 0, \\ \Big(\prod_{i=1}^{n_{d,s}} \beta\theta_{d,s}^{-\beta} t_{d,s,i}^{\beta-1}\Big) \exp\Big(-(\tau_{d,s}/\theta_{d,s})^{\beta}\Big), & \text{if } n_{d,s} > 0, \end{cases}
\end{aligned}
\tag{10}
$$

- Likelihood function of a hierarchical JPLP, written at shift-level (s)

$$
\begin{aligned}
& L^{*}_{d,s}(\kappa, \beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s) \\
& = \begin{cases} \exp\Big(-\int_0^{\tau_{d,s}} \lambda_{\text{JPLP}}(u) du\Big), & \text{if } n_{d,s} = 0, \\ \Big(\prod_{i=1}^{n_{d,s}} \lambda_{\text{JPLP}}(t_{d,s,i})\Big) \exp\Big(-\int_0^{\tau_{d,s}} \lambda_{\text{JPLP}}(u) du\Big), & \text{if } n_{d,s} > 0, \end{cases}
\end{aligned}
\tag{11}
$$

A More Clever Way of Writing Likelihood Function of JPLP

Likelihood function of a hierarchical JPLP, rewritten

$$
\begin{aligned}
& L^*_{d,s,r}(\kappa, \beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_r) = \\
& \begin{cases}
\exp\Big(-\int_{a_{d,s,r-1}}^{a_{d,s,r}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s,r} = 0, \\
\Big(\prod_{i=1}^{n_{d,s,r}} \lambda_{\text{JPLP}}(t_{d,s,r,i})\Big) \exp\Big(-\int_{a_{d,s,r-1}}^{a_{d,s,r}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s,r} > 0,
\end{cases}
\end{aligned}
\tag{12}
$$

where $L^*_{d,s,r}$ depends on driver d, shift s, and segment r.

- This likelihood function is written at segment-level (r), instead of shift s level.

Joint Likelihood Function for PLP and JPLP

- The joint likelihood function for PLP is driver- and shift-based product:

$$L_{\text{PLP}} = \prod_{d=1}^{D} \prod_{s=1}^{S_d} L_{d,s}(\beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s). \tag{13}$$

- While the joint likelihood function for JPLP is driver-, shift-, and segment-based product:

$$L^*_{\text{JPLP}} = \prod_{d=1}^{D} \prod_{s=1}^{S_d} \prod_{r=1}^{R_{d,s}} L^*_{d,s,r}(\kappa, \beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_r), \tag{14}$$

  - Drivers $d = 1, 2, \ldots, D$,
  - Shifts $s = 1, 2, \ldots, S_d$,
  - Segments $r = 1, 2, \ldots, R_{d,s}$.

Four chains are applied to each estimation, with 1,000 warmup and 4,000 post-warmup iterations drawn from the posterior distributions using `Stan`[15].

A Simulation Study

- 1,000 simulations for three scenarios ($D = 10, 25, 50, 75, 100$):
  - Data from a PLP and estimated assuming a PLP (PLP),
  - Data from a JPLP and estimated assuming a JPLP (JPLP),
  - Data from a JPLP, but estimated assuming a PLP (PLP $\leftarrow$ JPLP).
- Data generating process:

$$\begin{aligned} x_1 &\sim \text{Normal}(1, 1^2),\ x_2 \sim \text{Gamma}(1, 1) \\ x_3 &\sim \text{Poisson}(2),\ \tau_{d,s} \sim \text{Normal}(10, 1.3^2) \end{aligned} \tag{15}$$

- Parameter setting:

$$\begin{aligned} \beta &= 1.2,\ \kappa = 0.8 \\ \mu_0 &= 0.2,\ \sigma_0 = 0.5, \\ \gamma_{01}, \gamma_{02}, \ldots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2) \\ \gamma_1 &= 1,\ \gamma_2 = 0.3,\ \gamma_3 = 0.2 \\ \theta_{d,s} &= \exp(\gamma_{0d} + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3). \end{aligned} \tag{16}$$

Data Sets, Outcomes and Predictors

- 496 commercial truck drivers
- Pings are aggregated into:
  - 64,860 shifts and 180,408 segments
- 8,386 kinematic SCEs:
  - 3,941 (47%) headway events
  - 3,576 (42.6%) hard brakes
  - 869 (10.4%) collision mitigation.
- Outcome: time to the SCEs since the start of shifts $t_{d,s,i}$
- Predictors:
  - driver demographics (age, gender, and race)
  - weather (visibility, precipitation intensity and probability)
  - speed mean and standard deviation.

Among the shifts with at least one SCEs (N=6,112), 21.3% of them (N=1,302) have at least two SCEs in one shift.

4 Aim 3

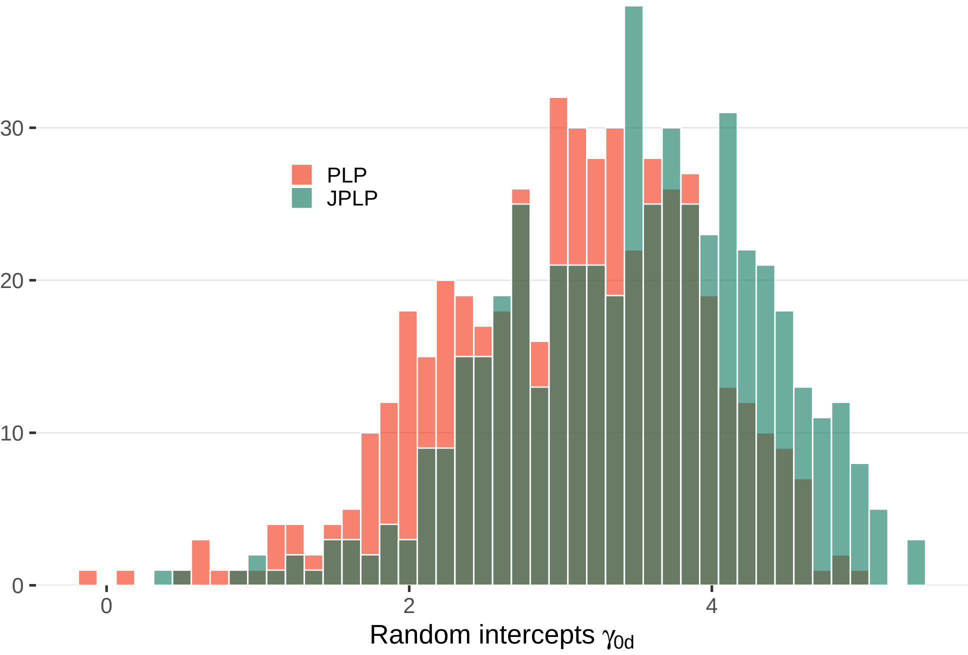
4.2 Results

Results for Simulation Study

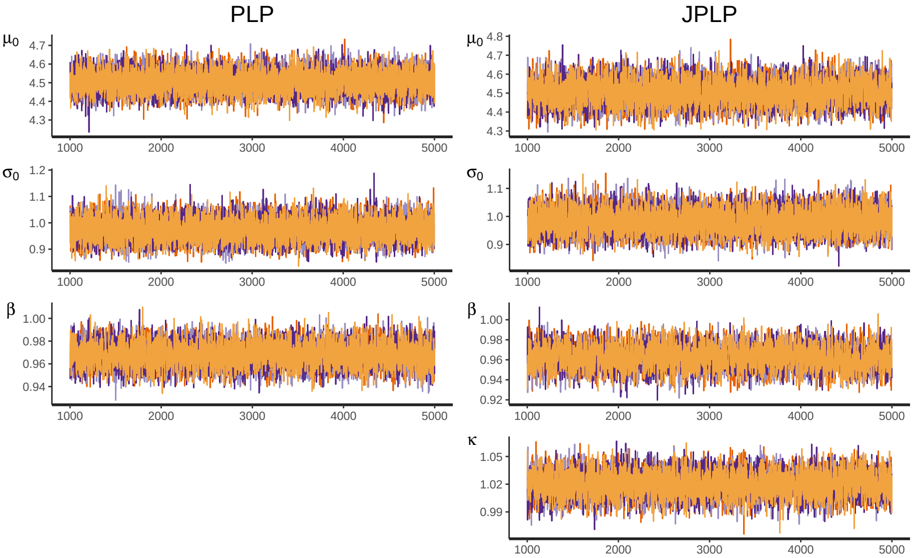
Scenario	D	estimate	β	κ	μ_0	σ_0	γ_1	γ_2	γ_3
PLP	10	bias Δ	-0.0102		-0.0282	0.0527	0.0203	0.0095	0.0067
PLP	25	bias Δ	-0.0045		-0.0015	0.0220	0.0066	0.0046	0.0012
PLP	50	bias Δ	-0.0017		-0.0068	0.0077	0.0040	0.0033	0.0005
PLP	75	bias Δ	-0.0017		-0.0026	0.0091	0.0034	0.0004	0.0007
PLP	100	bias Δ	-0.0006		-0.0034	0.0042	0.0009	0.0009	0.0003
PLP	10	S.E.	0.0589		0.2401	0.1722	0.0777	0.0696	0.0413
PLP	25	S.E.	0.0360		0.1392	0.0916	0.0459	0.0414	0.0247
PLP	50	S.E.	0.0254		0.0960	0.0610	0.0316	0.0286	0.0172
PLP	75	S.E.	0.0207		0.0784	0.0497	0.0258	0.0232	0.0139
PLP	100	S.E.	0.0179		0.0667	0.0420	0.0220	0.0198	0.0119
JPLP	10	bias Δ	-0.0226	0.0149	-0.0401	0.0696	0.0331	0.0218	0.0092
JPLP	25	bias Δ	-0.0131	0.0084	-0.0202	0.0219	0.0158	0.0081	0.0039
JPLP	50	bias Δ	-0.0057	0.0032	0.0014	0.0111	0.0037	0.0012	0.0039
JPLP	75	bias Δ	-0.0058	0.0028	0.0057	0.0097	0.0060	0.0012	0.0006
JPLP	100	bias Δ	-0.0043	0.0023	-0.0004	0.0041	0.0048	0.0003	0.0008
JPLP	10	S.E.	0.0828	0.0573	0.2556	0.1854	0.0992	0.0834	0.0498
JPLP	25	S.E.	0.0512	0.0360	0.1453	0.0960	0.0586	0.0477	0.0288
JPLP	50	S.E.	0.0366	0.0256	0.0999	0.0647	0.0406	0.0334	0.0201
JPLP	75	S.E.	0.0298	0.0208	0.0812	0.0519	0.0331	0.0272	0.0164
JPLP	100	S.E.	0.0258	0.0179	0.0699	0.0442	0.0287	0.0233	0.0141
PLP $\leftarrow$ JPLP	10	bias Δ	-0.1843		-0.1234	0.1599	0.1923	0.0645	0.0434
PLP $\leftarrow$ JPLP	25	bias Δ	-0.1740		-0.0866	0.1053	0.1769	0.0514	0.0374
PLP $\leftarrow$ JPLP	50	bias Δ	-0.1734		-0.0854	0.0977	0.1718	0.0531	0.0355
PLP $\leftarrow$ JPLP	75	bias Δ	-0.1724		-0.0874	0.0960	0.1686	0.0511	0.0346
PLP $\leftarrow$ JPLP	100	bias Δ	-0.1713		-0.0811	0.0925	0.1674	0.0512	0.0349
PLP $\leftarrow$ JPLP	10	S.E.	0.0580		0.2952	0.2078	0.1041	0.0946	0.0559
PLP $\leftarrow$ JPLP	25	S.E.	0.0354		0.1671	0.1095	0.0609	0.0546	0.0329
PLP $\leftarrow$ JPLP	50	S.E.	0.0250		0.1167	0.0743	0.0423	0.0383	0.0230
PLP $\leftarrow$ JPLP	75	S.E.	0.0204		0.0946	0.0601	0.0344	0.0310	0.0186
PLP $\leftarrow$ JPLP	100	S.E.	0.0177		0.0810	0.0514	0.0297	0.0266	0.0160

Results for JPLP and PLP

Parameters	Power law process				Jump power law process			
	mean	95% CI	$\hat{R}$	ESS	mean	95% CI	$\hat{R}$	ESS
$\hat{\beta}$	0.968	(0.948, 0.988)	1.000	6,500	0.962	(0.940, 0.985)	1.001	3,798
$\hat{\kappa}$					1.020	(0.995, 1.045)	1.000	5,400
$\hat{\mu}_0$	3.038	(2.397, 3.688)	1.001	2,979	3.490	(2.899, 4.091)	1.001	3,079
$\hat{\sigma}_0$	0.974	(0.897, 1.058)	1.000	9,581	0.982	(0.905, 1.066)	1.000	9,050
Age	0.003	(-0.005, 0.012)	1.001	2,250	0.004	(-0.005, 0.012)	1.001	2,566
Race: black	-0.113	(-0.329, 0.103)	1.002	1,951	-0.130	(-0.342, 0.087)	1.001	2,277
Race: other	-0.343	(-0.707, 0.021)	1.001	2,833	-0.361	(-0.729, 0.010)	1.001	3,334
Gender: female	-0.071	(-0.441, 0.300)	1.001	3,069	-0.071	(-0.435, 0.296)	1.001	4,162
Mean speed	0.019	(0.016, 0.023)	1.000	20,229	0.015	(0.013, 0.018)	1.000	19,827
Speed variation	0.026	(0.017, 0.034)	1.000	24,825	0.017	(0.013, 0.022)	1.000	13,127
Preci. intensity	-3.608	(-6.181, -0.935)	1.000	22,025	-2.136	(-3.785, -0.368)	1.000	24,397
Preci. prob.	0.397	(0.168, 0.628)	1.000	21,416	0.121	(-0.050, 0.296)	1.000	25,329
Wind speed	0.018	(0.008, 0.029)	1.000	32,980	0.010	(0.001, 0.018)	1.000	33,093


PLP
JPLP
30
20
10
0
0
2
4
Random intercepts γ_{0d}

Trace Plots for Selected Parameters



JPLP Estimation Results Stratified by SCE types

Parameters	Headway	Hard brake	Collision mitigation
$\hat{\beta}$	0.989 (0.956, 1.023)	0.922 (0.889, 0.955)	1.020 (0.950, 1.096)
$\hat{\kappa}$	1.034 (0.998, 1.071)	1.034 (0.996, 1.072)	0.890 (0.821, 0.964)
$\hat{\mu_0}$	7.096 (6.083, 8.139)	3.470 (2.770, 4.199)	4.729 (3.836, 5.666)
$\hat{\sigma_0}$	1.564 (1.411, 1.730)	1.073 (0.973, 1.182)	0.922 (0.786, 1.074)
Age	-0.006 (-0.020, 0.009)	0.011 (0.001, 0.021)	0.002 (-0.009, 0.012)
Race: Black	0.184 (-0.170, 0.546)	-0.312 (-0.565, -0.064)	0.113 (-0.153, 0.386)
Race: other	0.306 (-0.340, 0.967)	-0.539 (-0.968, -0.106)	0.100 (-0.373, 0.605)
Gender: female	0.266 (-0.343, 0.870)	-0.217 (-0.654, 0.230)	-0.181 (-0.675, 0.309)
Mean speed	-0.026 (-0.031, -0.021)	0.043 (0.039, 0.047)	0.039 (0.032, 0.046)
Speed variation	-0.009 (-0.017, -0.002)	0.017 (0.010, 0.024)	0.013 (-0.002, 0.027)
Preci. intensity	-0.771 (-4.306, 3.188)	-1.912 (-3.924, 0.269)	-0.676 (-6.329, 6.297)
Preci. prob.	0.694 (0.376, 1.015)	-0.495 (-0.724, -0.263)	0.808 (0.206, 1.423)
Wind speed	0.003 (-0.009, 0.015)	0.019 (0.005, 0.034)	0.000 (-0.025, 0.026)

5 Discussion

Summary of Findings and Contribution: Aim 1

- SCEs per 10,000 miles ↑
  - 8.4% (95% CI: 8-8.8%) crashes ↑,
  - 8.7% (95% CI: 4.8%-13.6%) injuries ↑.
- 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,
  - 22.2% (95% CI: 19.8-24.5%) for collision mitigation.
- Results are consistent in different business units and driver types.

Contribution to the field:

- Quantifies this association for commercial truck drivers
- Provide evidence for further data analytics studies.

Summary of Findings and Contribution: Aim 2

- Cumulative driving time is not associated with risk of SCEs,
- The null relationship is consistent across SCE types,
- A fair amount of variance explained by driver heterogeneity.

Contribution to the field:

- Proposes to use driver-centric instead of road-centric models,
- ping $\rightarrow$ shifts $\rightarrow$ trips $\rightarrow$ 30-minute interval framework.

Summary of Findings and Contribution: Aim 3

- Considerable amount of variability across drivers
- Different types of SCEs show different patterns:
  - Headway and hard brake tend to occur in early stages,
  - Collision mitigation tend to occur in later stages
- A considerable amount of variability across drivers.

<u>Contribution to the field:</u>

- Two driver-centric reliability models,
  - a Bayesian hierarchical NHPP with PLP intensity function,
  - a Bayesian hierarchical JPLP (a crucial κ parameter).
- The reliability patterns of different SCEs,
- R and Stan code for simulation and Bayesian estimation.

Limitations and Strengths

- Limitations
  - Exact time of the crashes were not recorded (SCEs $\rightarrow$ crash?)
  - Distances were approximated using the the haversine method
  - Traffic variables are not available
  - The assumption of proportion reliability jump may not hold
  - The length of breaks within shifts are ignored
- Strengths
  - Large sample size (>30,000 drivers and crashes)
  - Target commercial truck drivers
  - Investigation on human injuries and fatalities
  - Hierarchical models accounting for driver-level variance
  - Apply reliability models from reparable systems to truck driving.

Public Health Implications

- Safety-critical events are associated with crashes
- Driver-level variance needs to be accounted for in NDSs
- Driver-based ping data aggregation and analyses framework
- The importance of accounting for within-shift rests in reliability models

Future Research Directions

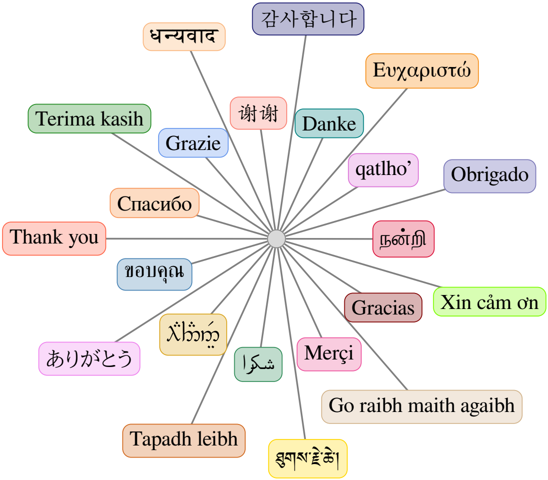
- SCEs need validation in other data sets using different sensors
- Similar framework in aim 2 and 3 validated among over-the-road drivers
- Accounting for the length of rest breaks, and the parameterization of reliability jump during break
- Subsampling MCMC algorithms for tall data.

Journal Publications & Submissions

1. **Cai, M.**, Alamdar Yazdi, M.A., Hu, Q., Mehdizadeh, A., Vinel, A., Davis, K.C., Xian, H., Megahed, F.M., Rigdon, S.E. “The association between crashes and safety-critical events: synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers”, 2nd-round revision at *Transportation Research Part C: Emerging Technologies*.
2. **Cai, M.**, Hu, Q., Mehdizadeh, A., Alamdar Yazdi, M.A., Vinel, A., Davis, K., Xian, H., Megahed, F., Rigdon, S.E. “Modeling safety-critical events using trucking naturalistic driving data: A driver-centric hierarchical framework for data analyses”, ready for submission to *Analytic Methods in Accident Research*.
3. **Cai, M.**, Mehdizadeh, A., Hu, Q., Alamdar Yazdi, M.A., Vinel, A., Davis, K.C., Xian, H., Megahed, F.M., Rigdon, S.E. “Hierarchical Point Process Models for Recurring Safety Critical Events for Commercial Truck Drivers”, submitted to *Journal of the American Statistical Association*.
4. Mehdizadeh, A., **Cai, M.**, Hu, Q., Alamdar Yazdi, M.A., Mohabbati-Kalejahi, N., Vinel, A., Rigdon, S.E., Davis, K.C. and Megahed, F.M. (2020) “A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling”, *Sensors*. DOI: 10.3390/s20041107, (co-first author).
5. Hu, Q., **Cai, M.**, Mohabbati-Kalejahi, N., Mehdizadeh, A., Alamdar Yazdi, M.A., Vinel, A., Rigdon, S.E., Davis, K.C. and Megahed, F.M. (2020) “A Review of Data Analytic Applications in Road Traffic Safety. Part 2: Prescriptive Modeling”, *Sensors*. DOI: 10.3390/s20041096.

Thank You!

- Steve Rigdon, PhD mentor
- Dissertation committee
- Collaborators in Auburn and Oxford (Amir, Qiong, . . .)
- PhD program directors & coordinators
- Colleagues at the VA
- Faculty members
- PhD cohort 2017 and fellow students
- My family in China
- All others



Questions?

miao.cai@outlook.com

References

[1] The White House. Remarks by President Trump celebrating America's truckers. https://www.whitehouse.gov/briefings-statements/remarks-president-trump-celebrating-americas-truckers/, 2020. [Issued on April 16, 2020; accessed July 03, 2020].

[2] Steven John. 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, 2019. [Published online June 3, 2019; accessed July 03, 2020].

[3] FMCSA. Large Truck and Bus Crash Facts 2016. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/safety/data-and-statistics/398686/ltbcf-2016-final-508c-may-2018.pdf, 2018. [Online; accessed 20-February-2019].

[4] Grant W Neeley and Lilliard E Richardson Jr. The effect of state regulations on truck-crash fatalities. *American Journal of Public Health*, 99(3):408–415, 2009.

[5] Eduard Zaloshnja, Ted Miller, et al. Unit costs of medium and heavy truck crashes. Technical report, The United States. Federal Motor Carrier Safety Administration, 2008.

[6] Amir Mehdizadeh, Miao Cai, Qiong Hu, Alamdar Yazdi, Mohammad Ali, Nasrin Mohabbati-Kalejahi, Alexander Vinel, Steven E Rigdon, Karen C Davis, and Fadel M Megahed. A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling. *Sensors*, 20(4):1107, 2020.

References (cont'd)

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

[8] Hal S Stern, Daniel Blower, Michael L Cohen, Charles A Czeisler, David F Dinges, Joel B Greenhouse, Feng Guo, Richard J Hanowski, Natalie P Hartenbaum, Gerald P Krueger, et al. Data and methods for studying commercial motor vehicle driver fatigue, highway safety and long-term driver health. *Accident Analysis & Prevention*, 126:37–42, 2019.

[9] Feng Guo. Statistical Methods for Naturalistic Driving Studies. *Annual Review of Statistics and Its Application*, 6:309–328, 2019.

[10] Thomas A Dingus, Richard J Hanowski, and Sheila G Klauer. Estimating crash risk. *Ergonomics in Design*, 19(4):8–12, 2011.

[11] Joseph H Saleh, Elizabeth A Saltmarsh, Francesca M Favaro, and Loic Brevault. Accident precursors, near misses, and warning signs: critical review and formal definitions within the framework of discrete event systems. *Reliability Engineering & System Safety*, 114:148–154, 2013.

[12] Anurag Pande, Sai Chand, Neeraj Saxena, Vinayak Dixit, James Loy, Brian Wolshon, and Joshua D Kent. A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accident Analysis & Prevention*, 101:107–116, 2017.

[13] Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2018. R package version 2.17.4.

References (cont'd)

[14] Kristina P Vatcheva, MinJae Lee, Joseph B McCormick, and Mohammad H Rahbar. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*, 6(2), 2016.

[15] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017.