



Contents lists available at ScienceDirect

# Accident Analysis and Prevention

journal homepage: [www.elsevier.com/locate/aap](http://www.elsevier.com/locate/aap)

## Driving safety assessment for ride-hailing drivers

Huiying Mao<sup>a,b,1</sup>, Xinwei Deng<sup>a</sup>, Honggang Jiang<sup>b</sup>, Liang Shi<sup>a,b,1</sup>, Hao Li<sup>b</sup>, Liheng Tuo<sup>b</sup>, Donghai Shi<sup>b</sup>, Feng Guo<sup>a,c,\*</sup>

<sup>a</sup> Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

<sup>b</sup> Didi Chuxing Technology Co., Beijing, China

<sup>c</sup> Virginia Tech Transportation Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

### ARTICLE INFO

#### Keywords:

Ride-hailing drivers  
Crash risk factors  
Operational characteristics  
General additive models  
SHapley Additive exPlanation

### ABSTRACT

Ride-hailing services, which have become increasingly prevalent in the last decade, provide an efficient travel mode by matching drivers and travelers via smartphone apps. Ride-hailing services enable millions of non-traditional taxi drivers to provide travel services, but may also raise safety concerns due to heterogeneity in the driver population. This study evaluated crash risk factors for ride-hailing drivers, including driving history and ride-hailing operational characteristics, using a sample of 189,815 drivers. We utilized the Poisson generalized additive model to accommodate for the potential nonlinear relationship between crash rate and risk factors. Results showed that crash history, the percentage of long-shift bookings, driving distance, operations during peak hours, years of being a ride-hailing driver, and passenger rating were significantly associated with crash risk. Several factors showed nonlinear relationships with crash risk. We adopted the SHapley Additive exPlanation (SHAP) method to assess and visualize the impact of each risk factor. The results indicated that passenger average rating, total driving distance, and crash history were the leading contributing factors. The findings of this study provide critical information for the development of safety countermeasures, driver education programs, as well as safety regulations for the ride-hailing industry.

### 1. Introduction

Ride-hailing services have grown exponentially in the last decade and have become a major component of the modern travel service industry. During this time, drivers working for ride-hailing services also increased substantially and now account for a considerable proportion of the overall driver population. This increase in riding-hailing traffic volume and driver population may lead to an increased number of crashes related to ride-hailing services. However, limited research has been conducted to understand the safety issues related to ride-hailing, and there is an urgent need to evaluate the safety of ride-hailing drivers and to identify the risk factors contributing to automobile crashes involving ride-hailing services.

Ride-hailing services provide peer-to-peer travel arrangements, which can be either for-profit or nonprofit. A smartphone app-based online platform connects passengers with ride-hailing drivers. Since Uber started its ride-hailing service in 2009, the sector has experienced significant growth in the United States and worldwide. Large companies like Didi Chuxing can host tens of millions of drivers and serve tens of

billions of trips annually. The increasing importance of ride-hailing services has inspired research into its impact on traffic congestion, total vehicle miles traveled (VMT), vehicle ownership, safety, and regulation policies.

Research shows mixed results with regard to ride-hailing services' impact on the total VMT as well on traffic crashes. Some studies have shown that ride-hailing services reduce total VMT, as many trips share similar origins and destinations, and ride-hailing services can combine multiple trips into one (Santi et al., 2014; Cici et al., 2014; Alexander and González, 2015; Alonso-Mora et al., 2017; Agatz et al., 2011). Conversely, there are also studies indicating an increase in VMT, as a large percentage of ride-hailing trips would otherwise have been made by walking, biking, or public transportation (Rayle et al., 2014; Schaller, 2017; Henao, 2017; Circella et al., 2018). In addition, extra travel due to passenger pick-up and drop-off, detouring, and cruising while waiting for ride bookings will also increase the total VMT (Schaller, 2017; Li et al., 2016). In terms of safety impacts, Barrios et al. (2018) found a 3% increase in the number of crashes by modeling crashes as a function of VMT and average driver quality. Dills and

\* Corresponding author at: Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA.

E-mail address: [feng.guo@vt.edu](mailto:feng.guo@vt.edu) (F. Guo).

<sup>1</sup> Dr. Mao and Mr. Shi worked as an intern in Didi Chuxing.

<https://doi.org/10.1016/j.aap.2020.105574>

Received 25 October 2019; Received in revised form 21 April 2020; Accepted 21 April 2020

0001-4575/© 2020 Elsevier Ltd. All rights reserved.

Mulholland (2018), on the other hand, showed a 17–40% decrease in fatal crashes for U.S. counties where Uber had operated for four or more years.

While the majority of the safety research in this area focuses on the impacts of ride-hailing at the societal level, it is also important to assess the driving risk among ride-hailing drivers. Unlike traditional professional taxi drivers, a large portion of ride-hailing drivers perform this work as a part-time job. They typically have not obtained rigorous training and screening, as is required by taxi or truck fleets. Considerable heterogeneity exists among drivers in terms of driving experience, working preference, and driver behavior, which could lead to drastic differences in individual crash risk.

Understanding ride-hailing drivers' safety and contributing risk factors is especially valuable since ride-hailing companies can use appropriate interventions or educational countermeasures to improve safety. Several attributes are unique to ride-hailing drivers and may be used to estimate individual driver risk. For example, ride-hailing operation requires interacting with a smartphone app, which has been well-established as related to crash risk (Redelmeier and Tibshirani, 1997; Klauer et al., 2014; Dingus et al., 2016; Guo et al., 2017, 2019). This higher risk associated with driver cellphone use has been established both at the individual driver and trip level (Fitch et al., 2013; Atwood et al., 2018; Farmer et al., 2015). Higher cellphone exposure among ride-hailing drivers therefore puts them at higher risk compared to regular drivers.

Certain ride-hailing drivers may choose to work more during peak hours when the demand is high, while others might choose to work more during off-peak hours to avoid congestion. As congested traffic typically imposes higher crash risk, operating during peak hours can be correlated with crash risk. Ride-hailing services typically offer passengers an option to rate the driver after a trip. The rating represents riders' satisfaction level, which reflects the driver's working attitude as well as their driving behavior. Passenger ratings thus could be a predictor for crash risk.

Driving fatigue is a major contributing factor to crash risk (Stern et al., 2019; Liu et al., 2019; Liu and Guo, 2019). Professional drivers are subject to strict regulations. The Hours of Service rules, set forth by the Federal Motor Carrier Safety Administration puts a limit on how many hours a commercial truck driver can drive in one working shift and the minimum number of off-duty breaks between two working shifts. Studies have shown that Hours of Services rules can help drivers get more sleep, thus mitigating the effect of fatigue (Hanowski et al., 2007; Banks, 2007). As a considerable number of ride-hailing drivers work for extended hours, there is a need to study the prevalence of driving long-shifts as well as what impact that has on safety.

Generalized linear models (GLMs), especially Poisson and negative binomial (NB) regression, are often used to assess driver risk and identify risk factors (Guo and Fang, 2013; Guo et al., 2015; Chen et al., 2016; Antin et al., 2017; Guo, 2019). Poisson and NB regression models assume the logarithm of crash rate to be a linear combination of the covariates, which is often not satisfied in practice. For example, it is well known that age has a bathtub-shaped relationship with crash risk; i.e., young and senior drivers have higher crash rates than middle-aged drivers. Generalized additive models (GAM) can accommodate nonlinearities smoothly using proper basis functions (Hastie and Tibshirani, 1987). While the GAM is a common method in many disciplines, only a very limited number of traffic safety research studies have adopted GAM (Friedman et al., 2001; Zhang et al., 2012).

Quantitatively evaluating the impact of a contributing factor is crucial for risk assessment models. The SHapley Additive exPlanation (SHAP) method provides a unified approach for interpreting model outputs. The SHAP value measures the impact of each factor for the model outputs, for both individual observations and the study population. The additive property of SHAP assures that the summation of all the importance measures and baseline value adds up to the final output (Lundberg and Lee, 2017). A force-plot based on SHAP output can

provide a visualization of the impact for individual observations (Lundberg et al., 2018).

The objective of this paper is to evaluate the risk factors associated with ride-hailing drivers. The data include 189,815 active drivers from the Didi Chuxing Technology Corporation. A cross-sectional study design was used to examine risk factors associated with the crash risk in a 6-month period. Extensive data mining was conducted to extract features from operational characteristics. A Poisson GAM was used for model development and SHAP was used for the impact assessment of potential risk factors.

## 2. Material and methods

The ride-hailing data from Didi Chuxing are introduced in Section 2.1. The Poisson GAM, which can accommodate the potential nonlinear relationship, is presented in Section 2.2. We adopted the SHAP method to assess the contribution of risk factors to the fitted GAM as introduced in Section 2.3.

### 2.1. Ride-hailing data

The ride-hailing data were provided by the Didi Chuxing Technology Corporation. A cross-sectional design was adopted to include crash and driver information in the second half of the year 2018. The study population was defined as all active drivers from a major city in China. The study population was defined as all active drivers from a major city in China during the study period. The study sample included 189,815 drivers who completed more than 100 bookings during the second half of 2018. These 189,815 drivers combined drove two billion vehicle-kilometers during the study period and were involved in 5298 crashes. The average crash rate in our sample was 267.1 crashes per 100 million vehicle kilometers traveled.

The crashes included in the study occurred while the ride-hailing drivers were on-duty. The crashes could be reported by passengers, the ride-hailing drivers themselves, or the police. Since the crash reporting scheme differs from the police-reportable crash standard, the data include a considerable proportion of less severe crashes compared to typical police crash databases. This study includes all crashes with severity above the property damage level, including property damage only, injury, and fatal crashes.

Extensive data mining was conducted to extract ride-hailing operational features that might affect risk, primarily based on booking information. A booking is considered a completed trip from a passenger making a request online to the ride-hailing driver dropping off the passenger(s) at the destination. After a booking was completed, the passenger would either rate the ride-hailing driver on a five-star scale based on the service experience, or take no rating action. The booking length, passenger rating, start and end time, along with many other characteristics, were recorded in the booking database.

The vast operational characteristics data generated on the ride-hailing platform provide opportunities to study potential risk factors and quantify their effect on crash risk. For example, one feature is whether a driver took bookings during morning or evening peak hours, which are defined as 7–10 a.m. and 5–8 p.m., respectively. The complex traffic conditions during peak hours might impose a higher risk for drivers. Operational characteristics were aggregated from the detailed bookings database to the driver-level. Seven operational features were extracted at the driver level, as listed in Table 1.

Past crash history has been shown to be related to future crash risk and is used extensively in insurance and fleet safety management. A crash history variable was constructed to represent whether a driver was involved in crashes during on-duty revenue-generating working time in the years 2016 and 2017. The crash history for drivers who started after 2018 are labeled as "unknown" and are treated as a separate category.

A total booking distance variable was constructed to represent the

**Table 1**  
Descriptive statistics of ride-hailing operational characteristics data.

Categorical variable	Levels	Frequency	Percentage
Worked during morning peak (7:00 a.m.–9:00 a.m.)	No	9251	4.9%
	Yes	180,564	95.1%
Worked during evening peak (5:00 p.m.–8:00 p.m.)	No	3366	1.8%
	Yes	186,449	98.2%
Crash history (involved in a crash before?)	No	120,089	63.3%
	Yes	611	0.3%
	Unknown	69,115	36.4%
Continuous variable	Range	Mean (std. dev.)	Missing pct.
Total booking distance (10 <sup>4</sup> km)	0.02–5.94	1.05 (0.91)	0%
Years being ride-hailing drivers	0.30–4.60	2.28 (1.16)	0%
Average rating	1–5	4.92 (0.11)	0.9%
Percentage of long-shifts	0–1	0.33 (0.29)	38.3%

total working load and exposure to driving. The total booking distance is the cumulative revenue-generating driving distance of all the bookings a ride-hailing driver took during the study period. The total booking distance had dual roles in driver risk modeling. (1) The distance served as an exposure measure; i.e., drivers who drove more would have more chances to experience crashes due to longer exposure to at-risk conditions. (2) Drivers who drove different mileages could have different crash rates due to confounding factors, such as driving fatigue and driving experience. The analysis considered both aspects of the total booking distance variable.

Tenure duration can also be associated with crash risk in two ways. (1) A novice ride-hailing driver might not be familiar with the app interface and operational protocol, thus demanding a higher cognitive load, which could impair safe driving ability. (2) Tenure duration could also be associated with past safety performance; severe safety violations or crashes will lead to termination of the ride-share driving contract. Both familiarity with the operation and past safety performance are related to crash risk. We calculated the tenure history as the number of years the driver had worked for Didi Chuxing up until December 31, 2018.

Passengers' star ratings are a comprehensive rating of the quality of service, which includes driving behavior that could affect safety. The average passenger rating feature calculates the average star over all the bookings for which passengers gave a rating (1–5 stars) after a ride. Drivers who received less than three ratings were counted as missing. The ratings were missing for about 0.9% of the driver samples during

the second half of 2018 and these drivers were excluded from the analysis.

Driving over an extended period of time in a working shift could lead to driver fatigue and affect safety. We extracted a feature “percentage of long-shifts” to represent the tendency of a driver to work long-shifts. One data log point was recorded for every minute a ride-hailing driver was operating. By aggregating the total data points, segments of continuous online time were calculated for all trips. In addition, two consecutive online periods were considered to be one shift if the offline time between them was less than 7 h. A shift was considered a long-shift if the cumulative operating time was longer than 12 h. The percentage of long-shifts was calculated among all shifts for each driver in the study period.

The sheer volume of the log data made it difficult to retrieve achieved data from the study period. Instead, the online records from January 1, 2019, to March 31, 2019 were used. As the percentage of long-shifts is typically associated with a driver's working schedule and is relatively stable over time, we can assume the data from the first quarter of 2019 are representative of these driver characteristics. One issue is that a considerable number of ride-hailing drivers who operated during the second half of 2018 were missing from 2019, leading to a large portion of missing data. We adopted a method to deal with these missing values, as shown in Section 2.2, Eq. (4).

The Poisson and NB models assume a linear relationship between the risk factor and the logarithm of the crash rate, which is not satisfied by the data. Fig. 1 shows the relationship between average crash rate and the percentage of long-shifts, in which bars represent the percentage of drivers who fall into each long-shift category and the line represents the relationship between long-shift percentage and crash rate. As the figure shows, crash rate follows a pattern of a decreasing trend before 20–40%, followed by a sharp increase. The pattern of crash rate can be explained by the fact that inexperienced drivers usually do not take long-shifts and their crash risk is typically higher. Those who occasionally take long-shifts are more experienced drivers and tend to be more cautious while driving. Taking too many long-shifts will increase crash risk. Drivers who mainly take long-shifts are primarily professional drivers and their crash risk will therefore be lower. As the regular GLM cannot capture the nonlinear relationship as illustrated, the GAM is a more appropriate model for driver risk assessment.

## 2.2. Generalized additive model for count response

A regression model for count response was used to quantify the relationship between the continuous risk factors and the logarithm of

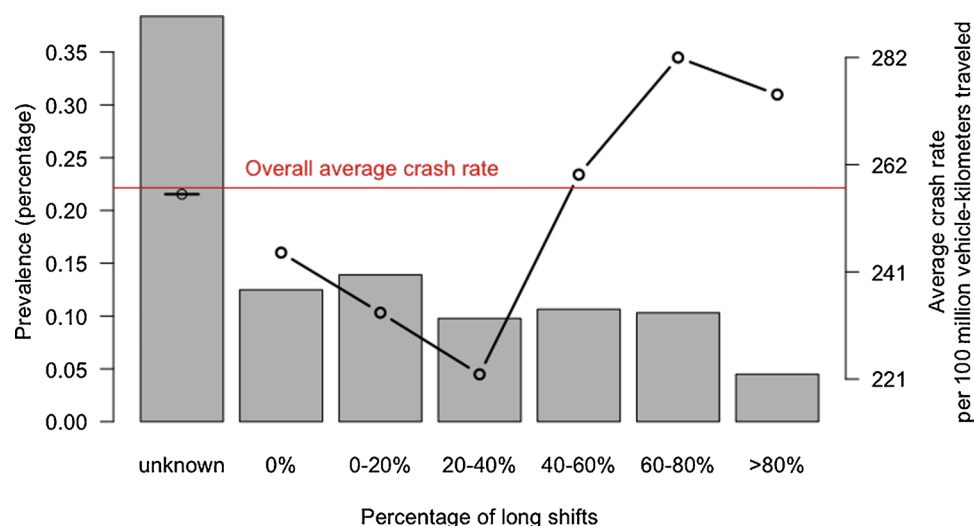


Fig. 1. The prevalence (bar) and average crash rate (segmented line) for variable ‘percentage of long-shifts’.

crash occurrence rate. Assume that the number of crashes for driver  $i$ ,  $Y_i$ , follows a Poisson or NB model; i.e.,

$$Y_i \sim \text{Poisson}(\lambda_i \cdot E_i) \text{ or } Y_i \sim \text{NB}(\lambda_i \cdot E_i, \gamma),$$

where  $E_i$  is the exposure measured by driving time or mileage;  $\lambda_i$  is the expected crash rate for driver  $i$ ; and  $\gamma$  is the overdispersion parameter for NB. Both Poisson and NB GAMs assume

$$\log(\lambda_i) = \beta_0 + f_1(X_{i1}) + f_2(X_{i2}) + \dots + f_p(X_{ip}),$$

$$i = 1, 2, \dots, n, \quad (1)$$

where  $X_{i1}, X_{i2}, \dots, X_{ip}$  are the  $p$  covariates for driver  $i$  used for risk assessment.  $f_j(\cdot)$  can be either a linear or nonlinear function in terms of the  $j$ th covariate,  $j = 1, \dots, p$ . When  $f_1(\cdot), \dots, f_p(\cdot)$  are all linear functions, a GAM degrades to a regular GLM.

The GAM uses a linear combination of proper basis functions to model a complex non-linear relationship (Agesti, 2003). Denote  $\{h_{j1}(\cdot), \dots, h_{jM_j}(\cdot)\}$  as the set of basis for function  $f_j(\cdot)$ ,  $j = 1, \dots, p$ . With basis expansion,  $f_j(\cdot)$  can be written as a linear combination of the basis:

$$f_j(\cdot) = \sum_{m=1}^{M_j} \beta_{jm} h_{jm}(\cdot), \quad j = 1, \dots, p. \quad (2)$$

With Eq. (2), the model in Eq. (1) can be viewed as a GLM in terms of

$$\underbrace{h_{11}(X_{i1}), \dots, h_{1M_1}(X_{i1})}_{\text{basis for } f_1(X_{i1})}; \dots; \underbrace{h_{p1}(X_{ip}), \dots, h_{pM_p}(X_{ip})}_{\text{basis for } f_p(X_{ip})}.$$

That is,

$$\log(\lambda_i) = \beta_0 + \sum_{m=1}^{M_1} \beta_{1m} h_{1m}(X_{i1}) + \dots + \sum_{m=1}^{M_p} \beta_{pm} h_{pm}(X_{ip}),$$

$$i = 1, 2, \dots, n. \quad (3)$$

We adopted the commonly used spline model with cubic basis functions. The knot selection was determined based on exploratory data analysis and domain knowledge. For some variables, such as “average rating,” we also adopted the smoothing spline for the construction of the basis functions. Note that the smoothing spline does not require knot selection but needs to specify the degree of freedom.

We also developed a strategy to deal with the missing value issue in the GAM. An indicator variable,  $\mathbb{I}(X_{ij} = \text{NA})$ , was used to represent the missing value for covariate  $X_{ij}$ . The GAM was extended to accommodate the missing value through the following additive form,

$$\beta_{i0} \cdot (X_{i1} = \text{NA}) + \sum_{m=1}^{M_1} \beta_{1m} (h_{1m}(X_{i1}) \cdot (X_{i1} \neq \text{NA}))$$

$$= \begin{cases} \beta_{i0}, & \text{if } X_{i1} = \text{NA} \\ \sum_{m=1}^{M_1} \beta_{1m} h_{1m}(X_{i1}), & \text{if } X_{i1} \neq \text{NA} \end{cases}$$

$$i = 1, 2, \dots, n. \quad (4)$$
  

$$= \begin{cases} \beta_{i0}, & \text{if } X_{i1} = \text{NA}, \\ \sum_{m=1}^{M_1} \beta_{1m} h_{1m}(X_{i1}), & \text{if } X_{i1} \neq \text{NA}, \end{cases} \quad i = 1, 2, \dots, n. \quad (5)$$

The coefficients  $\{\beta_0, \beta_{jm}; j = 1, \dots, p, m = 1, \dots, M_j\}$  can be estimated using a regular generalized linear regression fitting approach. The estimated crash rate for a driver  $i$  is

$$\log(\hat{\lambda}_i) = \hat{\beta}_0 + \sum_{m=1}^{M_1} \hat{\beta}_{1m} h_{1m}(X_{i1}) +$$

$$\hat{f}_1(X_{i1})$$

$$+ \hat{\beta}_{i0} \cdot \mathbb{I}(X_{ij} = \text{NA}) + \sum_{m=1}^{M_1} \hat{\beta}_{1m} (h_{1m}(X_{i1}) \cdot \mathbb{I}(X_{i1} \neq \text{NA})) + \dots$$

$$\hat{f}_1(X_{i1})$$

$$+ \sum_{m=1}^{M_p} \hat{\beta}_{pm} h_{pm}(X_{ip}), \quad i = 1, 2, \dots, n,$$

$$\hat{f}_p(X_{ip}) \quad (6)$$

$$+ \underbrace{\hat{\beta}_{i0} \cdot \mathbb{I}(X_{i1} = \text{NA}) + \sum_{m=1}^{M_1} \hat{\beta}_{1m} (h_{1m}(X_{i1}) \cdot \mathbb{I}(X_{i1} \neq \text{NA}))}_{\hat{f}_1(X_{i1})} + \dots \quad (7)$$

$$+ \underbrace{\sum_{m=1}^{M_p} \hat{\beta}_{pm} h_{pm}(X_{ip})}_{\hat{f}_p(X_{ip})}, \quad i = 1, 2, \dots, n, \quad (8)$$

where  $\{\hat{\beta}_0, \hat{\beta}_{jm}; j = 1, \dots, p, m = 1, \dots, M_j\}$  are the estimated coefficients. We used a cubic spline with two knots as an example for illustration. Suppose the knots used for  $f_j(\cdot)$  are  $\xi_1$  and  $\xi_2$ , then the following functions constitute the cubic spline basis:

$$h_{j1}(x) = x,$$

$$h_{j2}(x) = x^2,$$

$$h_{j3}(x) = x^3,$$

$$h_{j4}(x) = (x - \xi_1)_+^3,$$

$$h_{j5}(x) = (x - \xi_2)_+^3.$$

Then,

$$f_j(X_{ij}) = \beta_{j1} \cdot X_{ij} + \beta_{j2} \cdot X_{ij}^2 + \beta_{j3} \cdot X_{ij}^3 + \beta_{j4} \cdot (X_{ij} - \xi_1)_+^3 + \beta_{j5} \cdot (X_{ij} - \xi_2)_+^3. \quad (9)$$

The Poisson GAM introduced above can account for the nonlinear relationship between the logarithm of crash rate and the value of the risk factors. In the case that over-dispersion is present for a Poisson GAM, an NB GAM can be used to avoid the lack of fitting.

### 2.3. Model explanation using SHAP

Quantitatively assessing the impact of a risk factor to the crash rate both globally and for an individual driver is an important aspect of a prediction model. Global assessment of a risk factor refers to the impact at the driver population level, which provides crucial information for policy making and safety countermeasure development. At the individual driver level, the impact assessment can provide insight into why a driver is at higher risk and provide a tailored safety improvement plan for a specific driver.

The SHAP method provides a unified approach to explain the outputs of an arbitrary prediction model (Lundberg and Lee, 2017). The SHAP can be applied in a wide variety of models, including blackbox models, and is not affected by the unit of measurement. For generalized linear GAM models, the SHAP has an explicit form, as presented in this section. SHAP assigns each covariate an importance value for a particular prediction, both for population and for an individual observation. Denote  $\phi_{ij}$  as the importance of the  $j$ th risk factor to the model output for individual observation  $i$ . In the context of this paper, the model output is the logarithm of the estimated crash rate  $\log(\hat{\lambda}_i)$  for driver  $i$ . For linear models, the importance of the  $j$ th covariate for observation  $i$  has a simple form of:

$$\phi_{ij} = \hat{\beta}_j (X_{ij} - \frac{1}{n} \sum_{i=1}^n X_{ij}),$$

$$i = 1, \dots, n, \quad j = 1, \dots, p. \quad (10)$$

As GAM is linear in terms of the basis functions, the SHAP

**Table 2**  
Estimated model parameter for categorical variables.

	Estimate (CI)	CRR (CI)	P-value
Involved in crash(es): Y vs. N	0.497 (0.181, 0.813)	1.644 (1.198, 2.255)	0.0020*
Involved in crash(es): unknown vs. N	0.161 (0.079, 0.242)	1.175 (1.082, 1.274)	0.0001*
Morning peak: Y vs. N	0.493 (0.242, 0.745)	1.637 (1.274, 2.106)	0.0001*
Evening peak: Y vs. N	0.482 (0.050, 0.913)	1.619 (1.051, 2.492)	0.0286*

calculation follows a similar procedure. The importance value  $\phi_{ij}$  for GAM is

$$\phi_{ij} = \sum_{m=1}^{M_j} \hat{\beta}_{jm} [h_{jm}(X_{ij}) - h_{jm}(\bar{X}_j)],$$

$$i = 1, \dots, n, \quad j = 1, \dots, p, \quad (11)$$

where  $\{h_{j1}(\cdot), \dots, h_{jM_j}(\cdot)\}$  is the set of basis function for the  $j$ th covariate and

$$h_{jm}(\bar{X}_j) = \frac{1}{n} \sum_{i=1}^n h_{jm}(X_{ij}),$$

$$j = 1, \dots, p, \quad m = 1, \dots, M_j. \quad (12)$$

The additive property of SHAP guarantees that the summation of the importance measure along with the baseline value add up to the final model output, where the baseline value is the average model output, which is,

$$\begin{aligned} \log(\hat{\lambda}_i) &= \hat{\beta}_0 + \sum_{m=1}^{M_1} \hat{\beta}_{1m} h_{1m}(X_{i1}) + \dots + \sum_{m=1}^{M_p} \hat{\beta}_{pm} h_{pm}(X_{ip}) \\ &= \tilde{\beta}_0 + \sum_{m=1}^{M_1} \hat{\beta}_{1m} [h_{1m}(X_{i1}) - h_{1m}(\bar{X}_1)] + \dots + \sum_{m=1}^{M_p} \hat{\beta}_{pm} [h_{pm}(X_{ip}) \\ &\quad - h_{pm}(\bar{X}_p)] \\ &= \tilde{\beta}_0 + \phi_{i1} + \dots + \phi_{ip}, \end{aligned} \quad (13)$$

$$= \tilde{\beta}_0 + \sum_{m=1}^{M_1} \hat{\beta}_{1m} [h_{1m}(X_{i1}) - h_{1m}(\bar{X}_1)] + \dots + \sum_{m=1}^{M_p} \hat{\beta}_{pm} [h_{pm}(X_{ip}) - h_{pm}(\bar{X}_p)] \quad (14)$$

$$= \tilde{\beta}_0 + \phi_{i1} + \dots + \phi_{ip}, \quad (15)$$

where

$$\tilde{\beta}_0 = \hat{\beta}_0 + \sum_{m=1}^{M_1} \hat{\beta}_{1m} \cdot h_{1m}(\bar{X}_1) + \dots + \sum_{m=1}^{M_p} \hat{\beta}_{pm} \cdot h_{pm}(\bar{X}_p) \quad (16)$$

is the baseline value representing the average model output for all drivers and is a constant.

The SHAP can also be used to quantify the global impact of each risk factor, in which case the average absolute impact on model output magnitude is used as the evaluation metric; i.e.,

$$\sum_{i=1}^n |\phi_{ij}|/n, \quad j = 1, \dots, p. \quad (17)$$

The above global metric measures the impact on the logarithm of estimated crash risk over all drivers. This global measurement can be used to assess and compare the impact among multiple risk factors.

### 3. Results

We applied the GAM and SHAP modeling framework to the Didi Chuxing dataset to identify risk factors significantly associated with ride-hailing drivers' crash risk. The quantitative impacts of each factor in drivers' crash risk, both at the individual driver level and globally,

were evaluated by the SHAP method. Risk factors were ranked by their global impacts to the study's driver population. The impact of multiple risk factors to the crash rate of a single driver's risk was demonstrated using a force-plot, which visualizes both the direction and magnitude of the effects of the different factors.

#### 3.1. Model estimation results

All the continuous variables were included in GAM using spline methods. The cubic spline was used for modeling the effects of the continuous variables "total driving distance," "years being ride-hailing drivers," and "percentage of long-shifts." The smoothing spline was adopted for describing the effects of "average rating." The spline format and knots were selected so that the fitted effect matched the patterns of empirical average crash rate. The Poisson models showed no over-dispersion presence and no evidence of lack of fitting, and thus were adopted instead of the negative binomial models.

The estimated model coefficients for categorical variables are shown in Table 2. The crash rate ratio (CRR) is the exponential of the estimated regression coefficient, representing the ratio between the crash rate of two levels (e.g., with previous crashes or without). The Wald 95% confidence interval (CI) was used for interval estimation. Drivers with previous crashes had a significantly higher crash rate than drivers with a clean crash history (CRR, 1.644; 95% CI: [1.198, 2.255]). Drivers with an unknown crash history also showed a significantly higher crash rate compared to crash-free drivers (CRR: 1.175; 95% CI: [1.082–1.274]).

The business operational characteristics of drivers show considerable impacts on crash risk. Drivers who had business transactions during the morning peak hours (7–10 a.m.) were significantly riskier than those who did not (CRR, 1.637; 95% CI: [1.274–2.106]). Drivers who had business transactions during the evening peak hours (5–8 p.m.) also had a significantly higher crash rate than those who did not (CRR, 1.619; 95% CI: [1.051–2.492]).

Fig. 2 shows the effect of four continuous variables on the logarithm of the estimated crash rate; i.e., the corresponding  $f_j(\cdot)$  term in Eq. (1). The analysis of variance shows that all four continuous variables significantly contributed to the crash rate: average rating ( $F$ -statistic, 599.41;  $p$ -value < 0.001), total booking distance ( $F$ -statistic, 50.09;  $p$ -value < 0.001), years being ride-hailing drivers ( $F$ -statistic, 32.92;  $p$ -value < 0.001), and percentage of long-shifts ( $F$ -statistic, 34.64;  $p$ -value < 0.001).

The average passenger rating showed a strong nonlinear relationship with crash rate. When the average rating was between 4.5 and 5.0 stars, the crash rate decreased rapidly with the increase in star rating. However, there was essentially no difference in crash rate for drivers with an average rating between 3 and 4.5. This result implies that drivers with close to perfect ratings were substantially safer than those with ratings below 4.5, while there was no obvious difference for drivers with moderate to good ratings. The regular GLM failed to detect such nonlinear relationships and could lead to incorrect conclusions.

Fig. 2(b) and (c) show that the crash rate decreased with the increase in total booking distance and total years since the driver joined Didi Chuxing. The relationship monotonically decreased, indicating strong effects on crash risk. Fig. 2(d) shows the effect of the percentage of long-shifts on crash rate. The nonlinear curve indicates a complicated

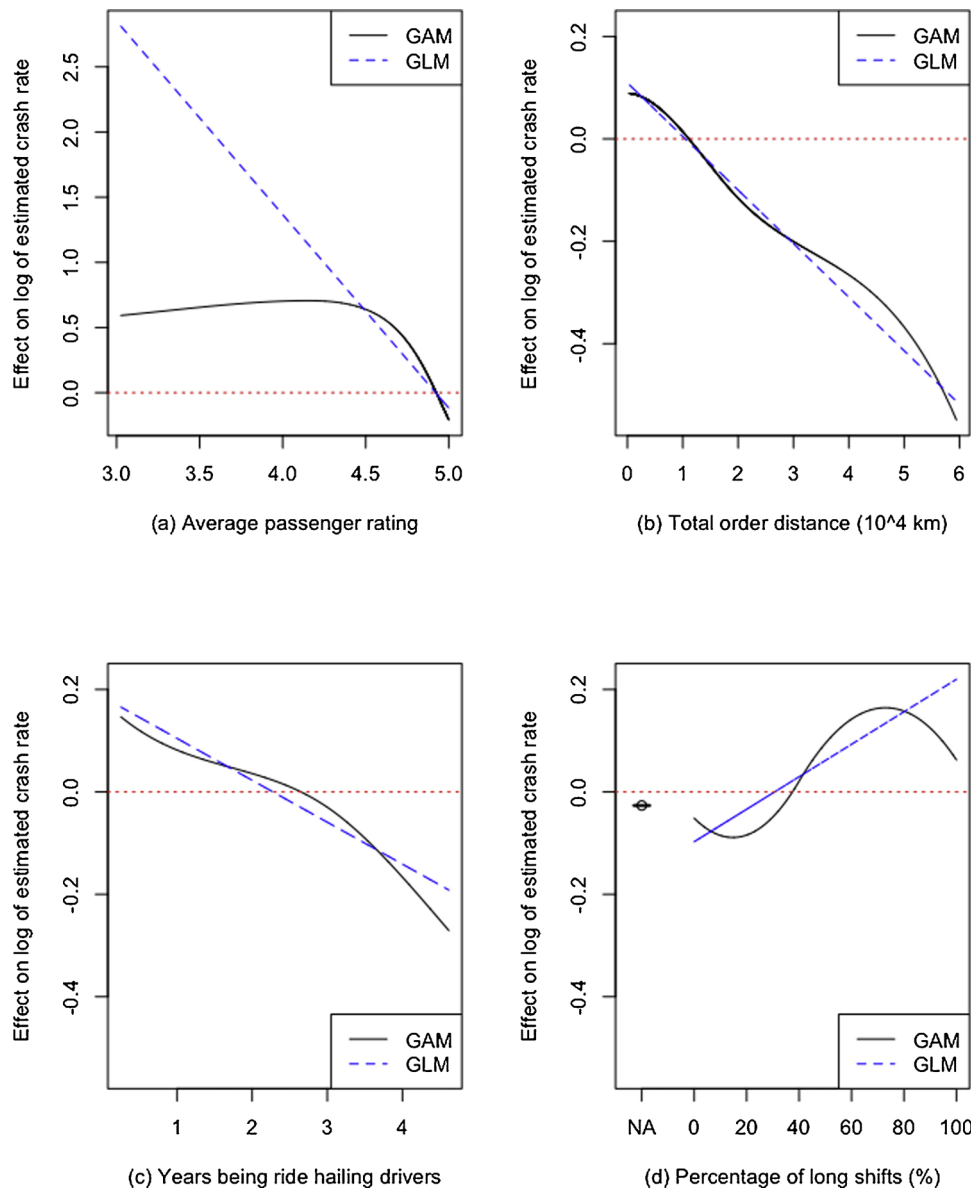


Fig. 2. GAM outputs: log crash rate versus risk factors; (a) average rating; (b) total driving distance; (c) tenure years; (d) percentage of long-shifts.

relationship. The results, in general, indicate that a higher percentage of long-shifts was associated with higher crash rate. However, the rate of change was highest when the percentage was between 20% and 70%. The crash rate change flattened out or even reversed at the lower and higher ends of the percentage range. This matches the empirical observation shown in Fig. 1.

### 3.2. Model explanation based on SHAP

We implemented the SHAP method to relate the impact of each risk factor to crash risk for both an individual driver and for the study driver population. The results are based on the estimated Poisson GAM presented in Section 3.1. The impact for an individual driver is visualized through a force-plot and the global impact is presented through a bar chart. The length of the bar on the left of the “output value” indicates the impact of a factor pushing the risk higher and vice versa. The force-plot clearly represents the risk of a single driver as well as the impact of each factor.

Fig. 3 shows the estimated model's force-plot for a randomly selected driver. The force-plot shows the point estimate of the outcome;

i.e., the logarithm of crash rate for an individual driver and the factors that have a negative and positive effect on the outcome. The “base value” indicates the average outcome for all drivers and the “output value” shows the outcome of the specific object, which in this context is the logarithm of this particular driver's estimated crash rate.

Fig. 3 shows that this particular driver is at higher risk than the average driver, as the output value is higher than the base value. Being involved in past crash is the biggest contributor to a higher crash rate. A lower average star rating and a higher percentage of long-shifts also makes this driver riskier than average. However, the risk is mitigated by the relative length of time (years) being a ride-hailing driver and the higher value of total booking distance.

Fig. 4 shows the ordered average impact of the seven risk factors used in the driver risk assessment model. The average impact shown in the x-axis is calculated as the average absolute SHAP value over all the drivers in the model using Eq. (17). As the figure shows, the average impact of a factor varies substantially. The passenger average rating has the largest impact, with a value almost twice as large as the second largest contributing factor, which is the total booking distance. The impact of crash history, years being a ride-hailing driver, and percent of

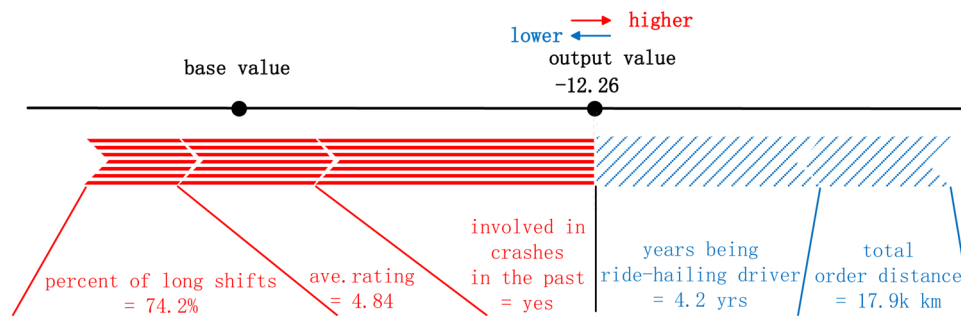


Fig. 3. SHAP force-plot of a single driver.

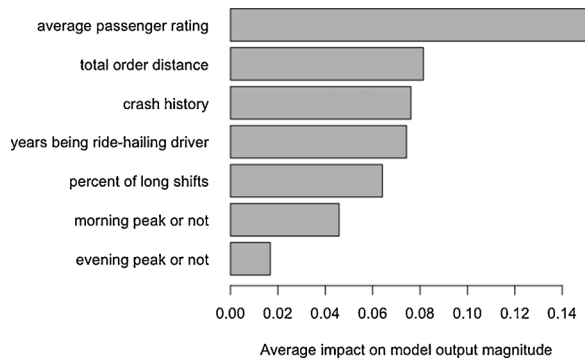


Fig. 4. The ordered impact of crash risk factors.

long-shifts are about the same. “Morning peak or not” has a larger impact than “evening peak or not.”

#### 4. Discussion and conclusion

Ride-hailing drivers constitute a unique and fast-growing driver population. Understanding the risk factors associated with this driver population is crucial for safety management programs as well as public policy making and regulation. This study used a large ride-hailing driver sample with hundreds of thousands of drivers and a large volume of ride-hailing operational data to quantitatively assess the impacts of potential risk factors.

In addition to traditional crash risk factors, ride-hailing drivers operate as semi-professional drivers and their resultant operational characteristics are shown to be associated with crash risk in this study. For example, drivers with near perfect passenger ratings are much safer than those with low ratings; drivers taking morning/evening peak bookings show significantly higher risk compared to those who do not; and the total booking distance contributes significantly to the estimation of crash risk. Finally, drivers who drove more had lower crash risk. These operational characteristics can help fleets develop appropriate driver safety management strategies.

Note that the current results for percent of long shifts relies on the assumption that drivers’ long-shift driving behavior remains similar from one time period to another. In future work, we will attempt to better align the period of log data with the study period to reduce the percentage of missing data to better reveal the underlying relationship.

This study expands the state-of-practice Poisson and NB GLM by using the GAM to incorporate the nonlinear relationships between the logarithm of crash rate and continuous risk factors. The results reveal the nonlinear risk profile by the percentage of long-shifts and average passenger rating. The total booking distance and tenure as a ride-hailing driver also show a certain level of nonlinear relationship with crash rate. In general, we recommend the use of GAM in traffic safety modeling, especially when the linear assumption is questionable based on exploratory analysis.

The SHAP provides an elegant way to decompose the model output into the impact of covariates. Traditional regression coefficients-based interpretation can only reflect the relative importance of a risk factor, providing virtually no information on the contribution to the model’s outcomes. The SHAP focuses on the predictive space and provides a quantitative assessment of the impact of both overall driver population as well as individual drivers. The force-plot visualizes the risk components for each driver and can facilitate risk management for individual drivers.

It should be noted that the time window for feature “percent of long shifts” was not aligned with the study period due to challenges in retrieving and processing vast amount of achieved ride-hailing data. Therefore, the results rely on the assumption that drivers’ long-shift driving behavior remains relative stable over time. In addition, we also developed a method to impute missing data caused by the mis-alignment. Early planning and data preparation can alleviate such challenges for future studies.

In summary, the big data collected by ride-hailing service companies provides unprecedented opportunities to evaluate driver risk and better understand the contributing risk factors for this unique driver population. The driving risk assessment framework described in this paper provides a novel modeling perspective and methods to interpret results. The finding of this study provide crucial information for driver fleet safety management programs, insurance, driver education programs, ride-hailing regulations, and public policy making.

#### Author statement

“Driving Safety Assessment for Ride-hailing Drivers”

This is a large research effort and the contribution of each author is listed below.

Huiying Mao and Liang Shi: Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Roles/Writing – original draft.

Xinwei Deng: Formal analysis; Investigation; Methodology; Supervision; Validation; Roles/Writing – original draft.

Honggang Jiang, Hao Li, and Donghai Shi: Conceptualization; Data curation; Investigation; Methodology; Resources; Supervision; Validation; Writing – review & editing.

Liheng Tuo: Conceptualization; Funding acquisition; Investigation; Project administration; Resources; Supervision; Validation; Writing – review & editing.

Feng Guo: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Roles/Writing – original draft.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was supported by a research grant from the XIAOJU Science and Technology Limited. We thank Yuxuan Gu, Chen Qian, Chenlin Wang, and Xinjian Peng for their help with the processing of driver characteristics and operational feature data.

## References

- Agatz, N., Erera, A.L., Savelsbergh, M.W., Wang, X., 2011. Dynamic ride-sharing: a simulation study in metro Atlanta. *Proc. Soc. Behav. Sci.* 17, 532–550.
- Agresti, A., 2003. *Categorical Data Analysis*, vol. 482. John Wiley & Sons, Hoboken.
- Alexander, L.P., González, M.C., 2015. Assessing the impact of real-time ridesharing on urban traffic using mobile phone data. *Proc. UrbComp* 1–9.
- Alonso-Mora, J., Samaranyake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc. Natl. Acad. Sci.* 114 (3), 462–467.
- Antin, J.F., Guo, F., Fang, Y., Dingus, T.A., Hankey, J.M., Perez, M.A., 2017. The influence of functional health on seniors' driving risk. *J. Transp. Health* 6, 237–244.
- Atwood, J., Guo, F., Fitch, G.M., Dingus, T.A., 2018. The driver-level crash risk associated with daily cellphone use and cellphone use while driving. *Accid. Anal. Prevent.*
- Banks, S., 2007. Behavioral and physiological consequences of sleep restriction. *J. Clin. Sleep Med.* 3 (5), 519–528.
- Barrios, J.M., Hochberg, Y.V., Yi, H., 2018. *The Cost of Convenience: Ridesharing and Traffic Fatalities*. University of Chicago Becker Friedman Institute for Economics Working Paper (2018-80).
- Chen, G.X., Fang, Y., Guo, F., Hanowski, R.J., 2016. The influence of daily sleep patterns of commercial truck drivers on driving performance. *Accid. Anal. Prevent.* 91, 55–63.
- Cici, B., Markopoulou, A., Frias-Martinez, E., Laoutaris, N., 2014. Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* 201–211.
- Circella, G., Alemi, F., Tiedeman, K., Handy, S., Mokhtarian, P., 2018. *The Adoption of Shared Mobility in California and its Relationship with Other Components of Travel Behavior*. Institute of Transportation Studies, University of California, Davis.
- Dills, A.K., Mulholland, S.E., 2018. Ride-sharing, fatal crashes, and crime. *South. Econ. J.* 84 (4), 965–991.
- Dingus, T.A., Guo, F., Lee, S.E., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J.M., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci.* 113 (10), 2636–2641.
- Farmer, C.M., Klauer, S.G., McClafferty, J.A., Guo, F., 2015. Relationship of near-crash/crash risk to time spent on a cell phone while driving. *Traffic Inj. Prevent.* 16 (8), 792–800.
- Fitch, G.M., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., Perez, M.A., Hanowski, R.J., Hankey, J.M., Dingus, T.A., 2013. *The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk Final Report*. Technical Report.
- Friedman, J., Hastie, T., Tibshirani, R., 2001. *The Elements of Statistical Learning*, vol. 1. Springer Series in Statistics, New York.
- Guo, F., 2019. Statistical methods for naturalistic driving studies. *Annu. Rev. Stat. Appl.* 6, 309–328.
- Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. *Accid. Anal. Prevent.* 61, 3–9.
- Guo, F., Fang, Y., Antin, J.F., 2015. Older driver fitness-to-drive evaluation using naturalistic driving data. *J. Saf. Res.* 54, 49–e29.
- Guo, F., Kim, I., Klauer, S.G., 2019. Semiparametric Bayesian models for evaluating time-varying driving risk factors using naturalistic driving data and case-crossover approach. *Stat. Med.* 38 (2), 160–174.
- Guo, F., Klauer, S.G., Fang, Y., Hankey, J.M., Antin, J.F., Perez, M.A., Lee, S.E., Dingus, T.A., 2017. The effects of age on crash risk associated with driver distraction. *Int. J. Epidemiol.* 46 (1), 258–265.
- Hanowski, R.J., Hickman, J., Fumero, M.C., Olson, R.L., Dingus, T.A., 2007. The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. *Accid. Anal. Prevent.* 39 (6), 1140–1145.
- Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. *J. Am. Stat. Assoc.* 82 (398), 371–386.
- Henao, A., 2017. *Impacts of Ridesourcing-Lyft and Uber on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior*. University of Colorado at Denver.
- Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted driving and risk of road crashes among novice and experienced drivers. *N. Engl. J. Med.* 46 (4), 600–601.
- Li, Z., Hong, Y., Zhang, Z., 2016. Do ride-sharing services affect traffic congestion? An empirical study of Uber entry. *Soc. Sci. Res. Netw.* 2002, 1–29.
- Liu, Y., Guo, F., 2019. A Bayesian time-varying coefficient model for multitype recurrent events. *J. Comput. Graph. Stat.* 0 (0), 1–12.
- Liu, Y., Guo, F., Hanowski, R.J., 2019. Assessing the impact of sleep time on truck driver performance using a recurrent event model. *Stat. Med.* 38 (21), 4096–4111.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 4765–4774.
- Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., et al., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* 2 (10), 749.
- Rayle, L., Shaheen, S.A., Chan, N., Dai, D., Cervero, R., 2014. *App-Based, On-Demand Ride Services: Comparing Taxi and Ridesourcing Trips and User Characteristics in San Francisco*. Citeseer Technical report.
- Redelmeier, D.A., Tibshirani, R.J., 1997. Association between cellular-telephone calls and motor vehicle collisions. *N. Engl. J. Med.* 336 (7), 453–458.
- Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. *Proc. Natl. Acad. Sci.* 111 (37), 13290–13294.
- Schaller, B., 2017. *Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City*. Schaller Consulting Technical Report.
- 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. *Accid. Anal. Prevent.* 126, 37–42.
- Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. *J. Saf. Res.* 43 (2), 107–114.