Truck Transportation Safety, Fatigue, Driver Characteristics and Weather: An Exploratory Data Analysis and Visualization

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

[a] *Saint Louis University, Saint Louis, MO, 63108*
[b] *Johns Hopkins University, Baltimore, MD, 21218*
[c] *Auburn University, Auburn, AL, 36849*
[d] *Miami University, Oxford, OH, 45056*

Abstract

This is the abstract.
It consists of two paragraphs.

1. Introduction

1.1. Background

Traffic safety is a pressing public health issue that involves huge lives losses and financial burden across the world and in the United States. As reported by the World Health Organization (WHO 2018b), road injury was the eighth cause of death globally in 2016, killing approximately 1.4 million people, which consisted of about 2.5% of all deaths in the world. If no sustained action is taken, road injuries were predicted to be the seventh leading cause of death across the world by 2030 (WHO 2018a). In the United States, transportation contributed to the highest number of fatal occupational injuries, leading to 2,077 deaths and accounting for over 40% of all fatal occupational injuries in 2017 (The United States, Bureau of Labor Statistics 2017). Traffic safety could also influence the economic growth of a country. Developing countries such as China and India could have suffered from 7-22% loss of per capita Gross Domestic Product over a 24-year period (Fumagalli et al. 2017).

Among all vehicles, large trucks are the primary concern of traffic safety since they are associated with more catastrophic accidents. In 2016, the Federal Motor Carrier Safety Administration (FMCSA) reported that 27% fatal crashes in work zones involved large trucks (FMCSA 2018). Among all 4,079 crashes involving large trucks or buses in 2016, 4,564 lives (1.12 lives per crash) were claimed in the accidents (FMCSA 2016). The economic losses associated with large truck crashes are also higher than those with passenger vehicles, with an estimated average cost of 91,000 US dollars per crash (Zaloshnja, Miller, and others 2008). The high risk of large trucks is attributed to two aspects of reasons (Huang et al. 2013). First, large truck drivers generally need to drive alone for long routes,

*Corresponding Author
Email addresses: miao.cai@slu.edu (Miao Cai), mza0052@auburn.edu (Mohammad Ali Alamdar Yazdi), qzh0011@auburn.edu (Qiong Hu), azm0127@auburn.edu (Amir Mehdizadeh), alexander.vinel@auburn.edu (Alexander Vinel), davisk4@miamioh.edu (Karen Davis), fmegahed@miamioh.edu (Fadel Megahed), steve.rigdon@slu.edu (Steven E. Rigdon)

under on-time demands, challenging weather and traffic conditions. On the other hand, trucks are huge weighted and potentially carrying hazardous cargoes.

To reduce the lives and economic losses associated with trucks, numerous studies attempted to screen the risk factors for truck-related traffic crashes or predict the crashes. The most common study design is a case-control study, matching a crash with one to up to ten non-crashes, and use statistical models such as logistic regressions to explain the causes or predict the crashes (Braver et al. 1997; Chen and Xie 2014; Meuleners et al. 2015; Née et al. 2019). This widespread case-control design is due to the fact that large truck crashes are very rare compared to the amount of time on road. However, a case-control study is limited in estimating the incidence data and may be contentious in selecting the control groups (Grimes and Schulz 2005; Sedgwick 2014).

Past truck safety literature almost exclusively focused on crashes, while ignoring the precursors to crashes. A precursor, or critical event, is a pattern or signature associated with an increasing chance of truck crash (Saleh et al. 2013; Janakiraman, Matthews, and Oza 2016). Truck critical events deserve more attention since they occur more frequently than crashes, suggest fatigue and a lapse in performance, and they can lead to giant crashes (Dingus et al. 2006). Although critcal events do not always result in an accident, they could be used as an early warning system to mitigate or prevent truck crashes (Kusano and Gabler 2012). describing the value of using real-time truck data and linking this background to quality since we are trying to submit to JQT.

1.2. Data Description and the Business Problem

1.2.1. Data source

The J.B. Hunt Transport Services, a trucking and transportation company in the United States, provided real-time ping data on 498 truck drivers who conducted regional work (REG, JBI00) from April 1st, 2015 to March 29th, 2016. A small device was installed in each truck in the company, which will ping irregularly (typically every 5-30 minutes). As Table 1 shows, each ping will collect data on the vehicle number, date and time, latitude, longitude, driver identity number, and speed at that second. In total, 13,187,289 pings were provided to the research team. Besides, the company also regularly collected real-time GPS location and time-stamped critical events data for all their trucks. There were 12,458 critical events occurred to these 498 truck drivers during the study period. Four types of critical events were recorded in this critical events data.

- Headway
- Hard brake
- Collision mitigation
- Rolling stability

We need a detailed description on how the JB Hunt define these critical events.

Apart from driver's characteristics and driving condition, weather also poses a threat on truck crashes and injuries (Zhu and Srinivasan 2011; Naik et al. 2016; Uddin and Huynh 2017). We obtained historic weather data

Table 1: A sample of ping data

TruckID	Driver1	Driver2	Date time	Speed	Latitude	Longitude
1	d1	d2	2015-11-16 13:56:00	9	33.42166	-84.14372
1	d1	d2	2015-11-16 13:57:00	0	33.41865	-84.14429
1	d1	d2	2015-11-16 13:57:00	0	33.41846	-84.14436
1	d1	d2	2015-11-16 13:59:00	28	33.41954	-84.14460
1	d1	d2	2015-11-16 14:05:00	2	33.41852	-84.14435

* The truck ID, driver1, driver2 have been anonymized to ensure privacy.

from the DarkSky Application Programming Interface (API), which allows us to query 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] and intensity, temperature, wind and others.

1.3. Research questions

We use exploratory data analysis, primarily data visualization, to address the following questions:

1. How are different measures of fatigue (cumulative driving time, rest time before a trip, rest time before a shift) associated with critical events?
2. Will time of driving influence the chances of critical events?
3. Are driver's characteristics (years of experience and age) associated with the chances of critical events
4. What is the relationship between weather conditions and critical events?

1.4. Innovation

- Assessing the precursors instead of crashes
- Real-time ping data
- Looking at all 498 drivers in a business section in a company instead of a road segment

2. Modeling Framework

2.1. Data preparation

- ping data
- critical events data
- driver's data
- weather data

[1] Ideally, the historic precipitation at a specific location and time should be yes or not. However, 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 some algorithms to infer the probability of precipitation in each location.

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 2 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 divded 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 Table 2 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.

Table 2: A sample of trips data

DriverID	TripID	Shift_ID	rest_time	start_time	end_time	Distance	trip_time	cumu_drive	CE_num	CE_type
1	1	1	10.50	2015-08-12T07:19:00Z	2015-08-12T12:47:00Z	279.45	5.47	5.47	2	1;1
1	2	1	1.40	2015-08-12T14:11:14Z	2015-08-12T14:22:14Z	7.11	0.18	5.65	0	
1	3	1	0.75	2015-08-12T15:07:40Z	2015-08-12T19:17:40Z	239.07	4.17	9.82	2	1;1
1	4	2	10.65	2015-08-13T05:57:34Z	2015-08-13T06:28:34Z	24.57	0.52	0.52	0	
1	5	2	0.37	2015-08-13T06:50:40Z	2015-08-13T11:03:40Z	233.22	4.22	4.73	1	1
1	6	2	1.18	2015-08-13T12:14:40Z	2015-08-13T12:40:40Z	22.00	0.43	5.17	2	1;1
1	7	2	1.22	2015-08-13T13:53:50Z	2015-08-13T17:05:50Z	133.05	3.20	8.37	3	1;1;1
1	8	2	0.40	2015-08-13T17:30:06Z	2015-08-13T20:08:06Z	93.37	2.63	11.00	1	1
1	9	3	10.18	2015-08-14T06:19:08Z	2015-08-14T06:44:08Z	7.67	0.42	0.42	1	1
1	10	3	0.38	2015-08-14T07:07:10Z	2015-08-14T10:47:10Z	188.72	3.67	4.08	0	
1	11	3	0.35	2015-08-14T11:08:48Z	2015-08-14T12:26:48Z	37.43	1.30	5.38	0	
1	12	3	0.75	2015-08-14T13:12:08Z	2015-08-14T18:02:08Z	204.50	4.83	10.22	2	1;1
1	13	4	10.43	2015-08-15T04:28:50Z	2015-08-15T10:11:50Z	234.92	5.72	5.72	1	1
1	14	4	49.37	2015-08-17T11:34:00Z	2015-08-17T12:42:00Z	31.56	1.13	6.85	0	
1	15	4	2.23	2015-08-17T14:56:36Z	2015-08-17T19:18:36Z	190.38	4.37	11.22	6	1;1;1;1;1;1
1	16	5	10.62	2015-08-18T09:19:16Z	2015-08-18T11:44:16Z	130.84	2.42	2.42	1	1
1	17	5	1.02	2015-08-18T12:45:40Z	2015-08-18T12:54:40Z	1.83	0.15	2.57	0	
1	18	5	0.95	2015-08-18T13:51:48Z	2015-08-18T15:17:48Z	75.75	1.43	4.00	0	
1	19	5	0.83	2015-08-18T16:07:52Z	2015-08-18T18:47:52Z	126.18	2.67	6.67	2	1;1
1	20	5	0.55	2015-08-18T19:21:12Z	2015-08-18T19:29:12Z	1.08	0.13	6.80	0	
1	21	6	10.92	2015-08-19T06:24:40Z	2015-08-19T11:36:40Z	214.04	5.20	5.20	0	
1	22	6	0.37	2015-08-19T11:59:30Z	2015-08-19T12:09:30Z	4.84	0.17	5.37	0	
1	23	6	0.73	2015-08-19T12:53:40Z	2015-08-19T19:27:40Z	222.78	6.57	11.93	3	1;1;1
1	24	7	10.78	2015-08-20T06:14:52Z	2015-08-20T12:29:52Z	299.36	6.25	6.25	1	1
1	25	7	0.72	2015-08-20T13:13:28Z	2015-08-20T16:00:28Z	146.06	2.78	9.03	1	1
1	26	8	10.48	2015-08-21T02:29:32Z	2015-08-21T05:03:32Z	80.74	2.57	2.57	0	
1	27	8	1.50	2015-08-21T06:34:12Z	2015-08-21T09:40:12Z	147.17	3.10	5.67	1	1
1	28	8	0.63	2015-08-21T10:19:10Z	2015-08-21T11:18:10Z	30.65	0.98	6.65	0	
1	29	8	2.05	2015-08-21T13:22:02Z	2015-08-21T13:29:02Z	3.39	0.12	6.77	0	
1	30	8	0.42	2015-08-21T13:54:20Z	2015-08-21T14:32:20Z	33.51	0.63	7.40	0	
1	31	9	10.85	2015-08-22T01:23:46Z	2015-08-22T04:39:46Z	144.60	3.27	3.27	2	1;1
1	32	9	0.58	2015-08-22T05:15:26Z	2015-08-22T07:38:26Z	85.45	2.38	5.65	1	3
1	33	9	1.88	2015-08-22T09:31:40Z	2015-08-22T11:57:40Z	116.32	2.43	8.08	0	
1	34	9	1.35	2015-08-22T13:19:12Z	2015-08-22T14:18:12Z	40.58	0.98	9.07	1	1

[*] CE denotes critical events.

[†] The truck ID, driver1, driver2 have been anonymized to ensure privacy.

2.2. *Exploratory data analysis*

All the data analyses and visualization were conducted in statistical computing environment `R` (R Core Team 2018). Specifically, data importing, cleaning and exporting were performed using the `data.table` and `dplyr` packages (Dowle and Srinivasan 2019; Wickham et al. 2018), the date and time objects were handled using the `lubridate` package (Grolemund and Wickham 2011), and all the visualizations were conducted using the `ggplot2` package (Wickham 2016).

3. Results

3.1. *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. (2018).

Driver's fatigue is difficult to measure in real life. In this study, we attempt to use three proxies to measure the fatigue of the truck drivers: cumulative driving time in a shift, the rest time before a shift, and the rest time before a trip.

3.1.1. *Cumulative driving time in a shift*

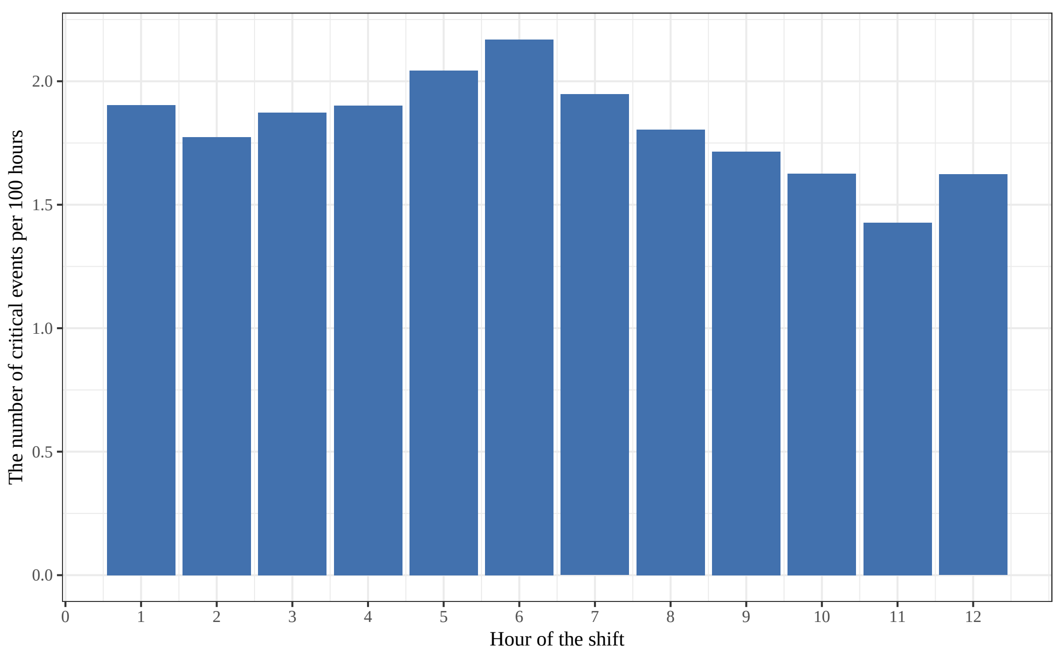

Figure 1: The number of critical events per 100 hours in each hour of the shift

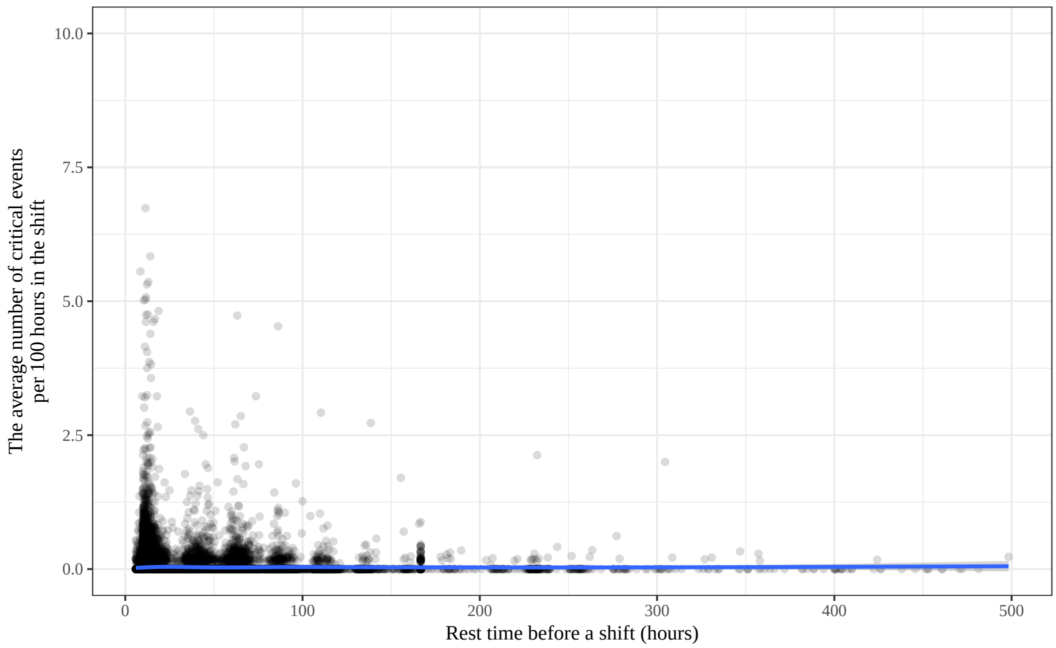

Figure 2: The number of critical events per 100 hours in a shift and the rest time before the shift

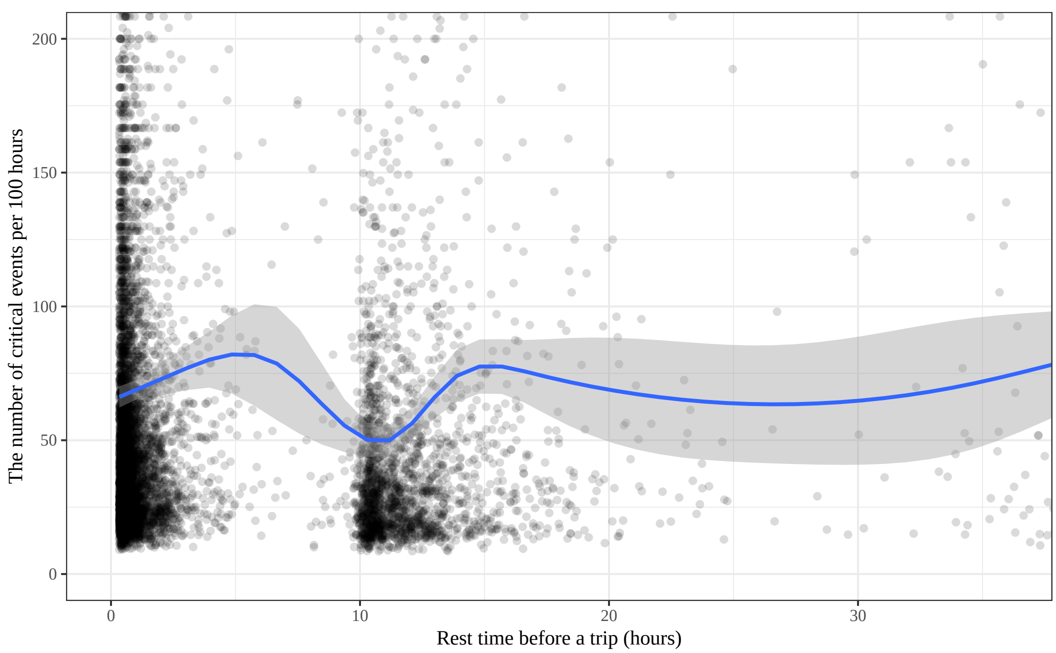

Figure 3: The number of critical events per 100 hours in a trip and the rest time before the trip

3.1.2. Rest time before a shift

3.1.3. Rest time before a trip

3.2. The time of driving

3.2.1. Day of the week

It was estimated that 84% of fatal crashes and 88% of nonfatal crashes related to large trucks occurred on weekdays (FMCSA 2018).

3.2.2. Hour of the day

FMCSA (2018) reported that 37%, 23% of injury-related crashes, and 20% of property damage only crashes associated with large trucks were from 6 p.m. to 6 a.m., which suggested that day or night may be a risk factor for crashes or critical events.

3.2.3. Sunrise and sunset times

During sunrise and sunset time, the visibility of the truck drivers are transitioning from clear to dark or the other way.

3.3. Driver characteristics

3.3.1. Age

The age of the truck driver could potentially influence their performance. Pinho et al. (2006) reported excessive sleepness among young truck drivers potentially caused by homeostatic pressure, poor sleep habits. De Craen et al. (2011). Gershon et al. (2019) reported that manual cellphone use and reaching for objects were associated with crash risk among teenage drivers.

3.3.2. Years of experience

3.4. Weather

3.4.1. Precipitation probability

3.4.2. Precipitation intensity

3.4.3. Visibility

4. Conclusion

4.1. Main contributions

4.2. Relevance to quality

4.3. Limitations

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, 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.

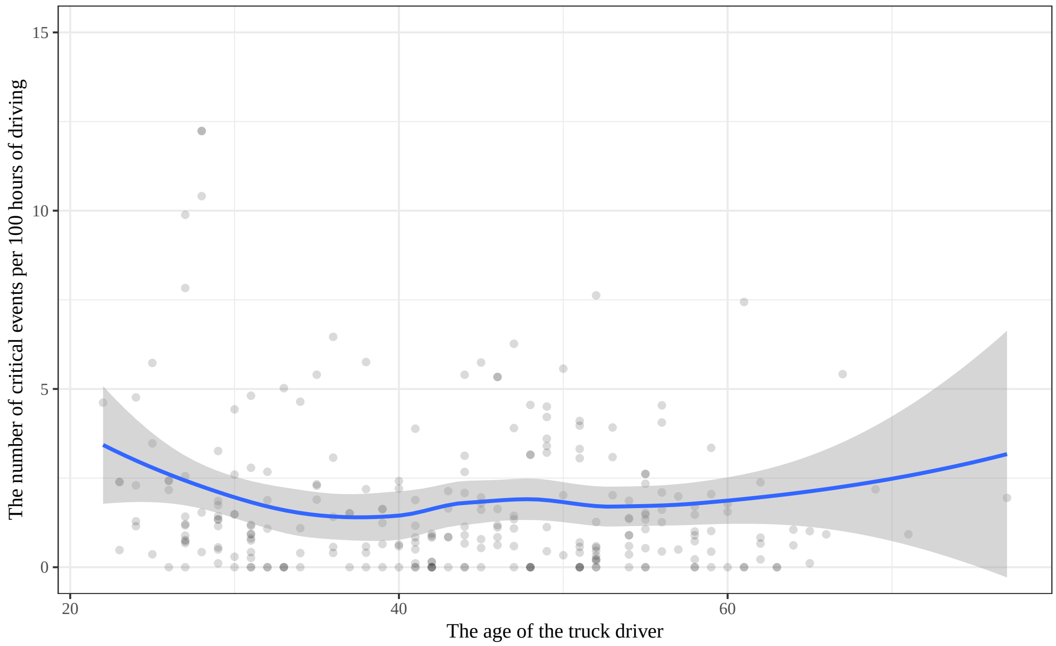

Figure 4: Truck driver's age and the rate of critical events

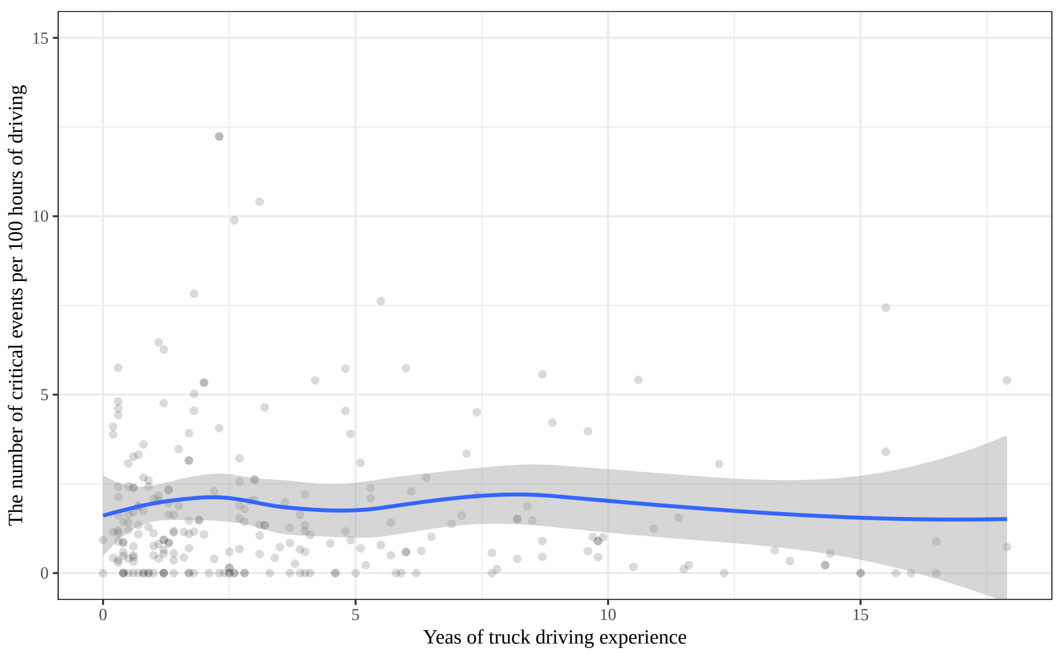

Figure 5: Truck driver's years of experience and the rate of critical events

Appendice

4.4. Sample data

4.5. R code

4.6. GitHub repository

References

Braver, Elisa R, Paul L Zador, Denise Thum, Eric L Mitter, Herbert M Baum, and Frank J Vilardo. 1997. "Tractor-Trailer Crashes in Indiana: A Case-Control Study of the Role of Truck Configuration." *Accident Analysis & Prevention* 29 (1). Elsevier: 79–96.

Chen, Chen, and Yuanchang Xie. 2014. "Modeling the Safety Impacts of Driving Hours and Rest Breaks on Truck Drivers Considering Time-Dependent Covariates." *Journal of Safety Research* 51. Elsevier: 57–63.

De Craen, S, DAM Twisk, Marjan Paula Hagenzieker, Hy Elffers, and Karel A Brookhuis. 2011. "Do Young Novice Drivers Overestimate Their Driving Skills More Than Experienced Drivers? Different Methods Lead to Different Conclusions." *Accident Analysis & Prevention* 43 (5). Elsevier: 1660–5.

Dingus, Thomas A, Vicki L Neale, Sheila G Klauer, Andrew D Petersen, and Robert J Carroll. 2006. "The Development of a Naturalistic Data Collection System to Perform Critical Incident Analysis: An Investigation of Safety and Fatigue Issues in Long-Haul Trucking." *Accident Analysis & Prevention* 38 (6). Elsevier: 1127–36.

Dowle, Matt, and Arun Srinivasan. 2019. *Data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

FMCSA. 2016. "Fatal occupational injuries by event, 2016." https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/safety/data-and-statistics/84856/cmvtrafficsafetyfactsheet2016-2017.pdf.

———. 2018. "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.

Fumagalli, E, Dipan Bose, Patricio Marquez, Lorenzo Rocco, Andrew Mirelman, Marc Suhrcke, and Alexander Irvin. 2017. "The High Toll of Traffic Injuries: Unacceptable and Preventable." World Bank.

Gershon, Pnina, Kellienne R Sita, Chunming Zhu, Johnathon P Ehsani, Sheila G Klauer, Tom A Dingus, and Bruce G Simons-Morton. 2019. "Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers." *American Journal of Preventive Medicine*. Elsevier.

Grimes, David A, and Kenneth F Schulz. 2005. "Compared to What? Finding Controls for Case-Control Studies." *The Lancet* 365 (9468). Elsevier: 1429–33.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. http://www.jstatsoft.org/v40/i03/.

Huang, Yueng-hsiang, Dov Zohar, Michelle M Robertson, Angela Garabet, Jin Lee, and Lauren A Murphy. 2013. "Development and Validation of Safety Climate Scales for Lone Workers Using Truck Drivers as Exemplar." *Transportation Research Part F: Traffic Psychology and Behaviour* 17. Elsevier: 5–19.

Janakiraman, Vijay Manikandan, Bryan Matthews, and Nikunj Oza. 2016. "Discovery of Precursors to Adverse Events Using Time Series Data." In *Proceedings of the 2016 Siam International Conference on Data Mining*, 639–47. SIAM.

Kusano, Kristofer D, and Hampton C Gabler. 2012. "Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions." *IEEE Transactions on Intelligent Transportation Systems* 13 (4). IEEE: 1546–55.

Meuleners, Lynn, Michelle L Fraser, Matthew H Govorko, and Mark R Stevenson. 2015. "Obstructive Sleep Apnea, Health-Related Factors, and Long Distance Heavy Vehicle Crashes in Western Australia: A Case Control Study." *Journal of Clinical Sleep Medicine* 11 (04). American Academy of Sleep Medicine: 413–18.

Naik, Bhaven, Li-Wei Tung, Shanshan Zhao, and Aemal J Khattak. 2016. "Weather Impacts on Single-Vehicle Truck Crash Injury Severity." *Journal of Safety Research* 58. Elsevier: 57–65.

Née, Mélanie, Benjamin Contrand, Ludivine Orriols, Cédric Gil-Jardiné, Cedric Galéra, and Emmanuel Lagarde. 2019. "Road Safety and Distraction, Results from a Responsibility Case-Control Study Among a Sample of Road Users Interviewed at the Emergency Room." *Accident Analysis & Prevention* 122. Elsevier: 19–24.

Pinho, Rachel SN de, Francisco P da Silva-Júnior, Joao Paulo C Bastos, Werllen S Maia, Marco Tulio de Mello, Veralice MS de Bruin, and Pedro Felipe C de Bruin. 2006. "Hypersomnolence and Accidents in Truck Drivers: A Cross-Sectional Study." *Chronobiology International* 23 (5). Taylor & Francis: 963–71.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Saleh, Joseph H, Elizabeth A Saltmarsh, Francesca M Favaro, and Loic Brevault. 2013. "Accident Precursors, Near Misses, and Warning Signs: Critical Review and Formal Definitions Within the Framework of Discrete Event Systems." *Reliability Engineering & System Safety* 114. Elsevier: 148–54.

Sedgwick, Philip. 2014. "Case-Control Studies: Advantages and Disadvantages." *Bmj* 348. British Medical Journal Publishing Group: f7707.

Stern, Hal S, Daniel Blower, Michael L Cohen, Charles A Czeisler, David F Dinges, Joel B Greenhouse, Feng Guo, et al. 2018. "Data and Methods for Studying Commercial Motor Vehicle Driver Fatigue, Highway Safety and Long-Term Driver Health." *Accident Analysis & Prevention*. Elsevier.

The Dark Sky Company, LLC. 2019. "Dark Sky API — Overview." https://darksky.net/dev/docs.

The United States, Bureau of Labor Statistics. 2017. "Fatal occupational injuries by event, 2017." https://www.bls.gov/charts/census-of-fatal-occupational-injuries/fatal-occupational-injuries-by-event-drilldown.htm.

Uddin, Majbah, and Nathan Huynh. 2017. "Truck-Involved Crashes Injury Severity Analysis for Different

Lighting Conditions on Rural and Urban Roadways." *Accident Analysis & Prevention* 108. Elsevier: 44–55.

WHO. 2018a. "Road traffic injuries." http://www.who.int/mediacentre/factsheets/fs358/en/.

———. 2018b. "The Top 10 Causes of Death." http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Zaloshnja, Eduard, Ted Miller, and others. 2008. "Unit Costs of Medium and Heavy Truck Crashes." The United States. Federal Motor Carrier Safety Administration.

Zhu, Xiaoyu, and Sivaramakrishnan Srinivasan. 2011. "A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes." *Accident Analysis & Prevention* 43 (1). Elsevier: 49–57.