# Response to Reviewers' Comments for TRC_2019_1394

We would like to thank the three anonymous reviewers for their comments, which significantly improved our manuscript. In this document, we detail how we have addressed all the comments made by the reviewers. For your convenience, we first reproduce your comments in a grey box , then highlight our response in **bold**. Changes in the manuscript are highlighted in blue, with the exception of minor edits on typos and language issues.

Remained issues:

1. Wait for J.B. Hunt on updates about the SCEs data description.
2. Add table 1 for descriptive statistics as requested.
3. Add correlation matrix and VIF test.
4. Steve, Faddel, Alex: 1) clarify Introduction, 2) clarify ping and SCEs data, 3) strengthen Discussion and Conclusion.

## 1 Response to Reviewer 1's Comments

### 1.1 General Comments for Part 1 and Part 2:

> The paper reads well and this study explores the association between truck crashes and safety-critical events using crash reports and naturalistic driving data. The topic is interesting. However, the reviewer has some concerns.
>
> 1. Justification of the research gap is weak, especially the introduction. Besides, what are the safety-critical events (SCEs)? Could you provide a definition?

**Thank you for your comments and suggestions. We have comprehensively revised the Introduction. The definition of safety-critical events (SCEs) has been given in the second paragraph on Page 3:**

Detected from dynamic kinematic 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 maneuver.

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

**Thank you for your suggestion. We have revised the Introduction substantially and we hope this version is clearer.**

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

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

How to choose the thresholds? Any references?

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

Need to fix these comments after we contact with J.B. Hunt.

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

**Thank you for your suggestion. We have added a table of the summary statistics for the predictor variables.**

5. Were all variables included in the models? So the authors did not consider the multicollinearity? And the tables made the reviewer confused.

Thn

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

**Yes. We added the following statement into the methods section.** Steve, please evaluate whether the following statement is correct or not.

A 95% credible interval is the probability that the parameter of interest falls into that range is 95%. If the 95% CI of the IRR includes one, we cannot say that the chance of IRR including one is very unlikely, so we will consider the parameter of this variable as statistically insignificant. On the other hand, if the 95% CI of the IRR does not include one, we will consider the parameter as statistically significant.

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

deeper analysis. Otherwise, the contributions of this manuscript would be limited.

Not quite sure what the reviewer means by deeper analysis?

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

Steve, Alex, and Fadel are probably more experienced than me in enhancing the discussion and conclusions.

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

For instance, page 11: One unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% CI: 8-8.8%) increase in the number of crashes per mile. How do you know 8.4%?

**Thank you for your comment. Here the description of our findings is written in a style that is more familiar in the field of epidemiology studies. A bit of math could reach the conclusion. Since the number of SCEs per mile** $\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K)$**, the rate change of** $\mu$ **when the predictor** $x_1$ **is increased by one unit can be calculated as:**

$$\frac{\mu'}{\mu} = \frac{\exp[\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \cdots + \beta_K x_K]}{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K} \tag{1}$$

$$= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K) \times \exp(\beta_1)}{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K)} \tag{2}$$

$$= \exp(\beta_1) \rightarrow \text{which is IRR.} \tag{3}$$

**Therefore, if the IRR is 1.084, the number of SCEs per mile** $\mu$ **is increased by 8.4% when the predictor** $x_1$ **is increased by one unit. To avoid the confusion in different columns of Table 2, we have added the model name and column number to the results:**

The pooled model (column 2 in Table 2) suggests that one unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% CI: 8-8.8%) increase in the number of crashes per mile.

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

**Thank you for your information. Here are the three links we provided:**

- website: https://caimiao0714.github.io/Github-SCE-crash/

- RMarkdown: https://github.com/caimiao0714/Github-SCE-crash/blob/master/index.Rmd
- data sets: https://github.com/caimiao0714/Github-SCE-crash/tree/master/data

**We have tested it on different computers and browsers and they work out well. Please check if there is any internet connection issues and try different web browsers and see if they work.**

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

**Thank you for your suggestion. We have corrected this spelling mistake, and we also conducted spelling and grammar check in this revised manuscript.**

## 2 Response to Reviewer 2's Comments

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

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

**Thank you for your suggestion. We have reduce the frequency of acronyms in our manuscript, and only commonly used acronyms are kept: naturalistic driving study (NDS), safety-critical events (SCEs), credible intervals (CIs).**

> 2. Please provide more details on the information about the "ping". For example,
>
> - Is a ping a single data point or several points?
> - The paper mentioned "ping data every couple of seconds to around 5 minutes." For a particular truck/driver, is the interval fixed or varying?
> - If a vehicle was not on, will a ping still be sent? Figure 1 seems implying so.
> - One important issue is that if a SCE occurred not during at ping period, will it

still show in the dataset? Some trigger based system will catch all such events, for example, when the acceleration is above 0.5G, a record will be automatically generated. It is not clear from the description whether this is the case for the data used.

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

- **One ping is a one data point including latitude, longitude, date, time, real-time speed, and others.**
- **For a particular truck/driver, the interval is still varying.**
- **Yes, the pings are still recorded when a vehicle was not on.**
- **The SCEs are collected in a system independently from the ping collecting system. These SCEs are collected whenever the kinematic thresholds are triggered by the driver. The definitions and kinematic thresholds were defined in Subsection 3.1 Data Description.**

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

Need to fix these comments after we contact with J.B. Hunt.

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

**Thank you for your suggestion. We have added a correlation matrix plot for the four different types of SCEs. We also conducted a variance inflation factor test for potential multi-collinearity. The Pearson correlation coefficients were all less than 0.2, and the VIF scores were all less than 1.5. The two tests suggest there were minor positive correlation**

**among the rates of four types of SCEs, but did not result in significant multi-collinearity problem.**

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

Add summary statistics for requested variables.

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

**Thank you for your correction. We have revised the sentence in the manuscript into the following one:**

NDS data are typically collected at a high frequency; for example, the acceleration and video data in SHRP2 NDS [1] and 100-Car NDS [2] were collected at a rate of 10 hertz from ignition-on to ignition-off. Therefore, the amount of NDS data are generally very large, and this provides both an opportunity and a challenge for data analytics.

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

Consider deleting the PPC figures.

Minor issues:

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

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

1. **A subscript "i" has been added to the ys in the right side of Equation (1).**
2. **The sentence has been changed into: "The large sample size provides a large statistical power to detect potential relationship between SCEs and crashes."**
3. **The corresponding sentence has been changed into "In NDSs, safety-critical events (SCEs) are commonly used to measure safety since crashes are very rare.".**
4. **PPC is graphical posterior predictive checks, which has been spelled out in the caption.**

## 3 Response to Reviewer 3's Comments

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

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

We could respond after we touch base with the J.B. Hunt guys. Another question is do we want to have another data paper for "Data in Brief". The overhead of writing this paper is very small. The APC charge is 600 USD.

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

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

**Thank you for your suggestion. We have added the references and updated Table 1.** Need to add the references in Table 1.

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

**Thank you for your suggestion. We have updated our description of the data ping, including the frequency of the data pings.**

Please explain the collision mitigation surrogate in table1.

Need to be done after we contact J.B. Hunt.

The authors should explain the method that they calculated headway. Were the vehicles instrumented with radar?

Need to be done after we contact J.B. Hunt.

More explanation of the hard brakes and the threshold that was used should be added. What was the threshold for the 231101 hard brakes? Clarify whether this number represents the events or data pings.

Need to be done after we contact J.B. Hunt.

**The number of SCEs (hard brakes, headway, collision mitigation, and rolling stability) represents the events data, which were collected independently from the ping data.**

Rolling stability should be defined.

Need to be done after we contact J.B. Hunt.

Description of the headway calculation and the threshold for critical headway SCE should be added.

Need to be done after we contact J.B. Hunt.

What are the present thresholds on page 8?

Need to be done after we contact J.B. Hunt.

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

**There are two references suggested by reviewer2 [3, 4, 5, 6].**

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

**Thank you for your suggestion. We have added more contents on missing values and data quality control in the first paragraph of Data Description subsection:**

The research team excluded 2,520 (7.4%) drivers who 1) had less than 100 pings, 2) cannot be matched in driver demographics table, and 3) were identified as obvious outliers regarding the rates of SCEs. In total, $1,494,678,173$ pings collected from 31,690 commercial truck drivers were included in the current study, and 98.7% of these pings were marked as having good quality. The latitudes and longitudes had at least five decimal places, which were worth up to 1.1 meters, indicating high-precision GPS location data.

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

**Thank you for your comment. As shown in Figure 2, this data set covers national general truck transporting environment in the United States. Although most of the ping data were in urban areas, rural areas were also covered.**

Xk should be xik

**Thank you for your suggestion. We have corrected that subscript issue.**

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

**Thank you for your comment. The Pareto** $k$ **values were not proposed by the authors, but by [7, 8]. We have added citations in the Results section.**

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

Add summary Table 1.

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

**Thank you for your suggestion. We have revised the manuscript accordingly.**

## References

[1] Jonathan M Hankey, Miguel A Perez, and Julie A McClafferty. Description of the shrp 2 naturalistic database and the crash, near-crash, and baseline data sets. Technical report, Virginia Tech Transportation Institute, 2016.

[2] Thomas A Dingus, Sheila G Klauer, Vicki Lewis Neale, Andy Petersen, Suzanne E Lee, Jeremy Sudweeks, Miguel A Perez, Jonathan Hankey, David Ramsey, Santosh Gupta,

et al. The 100-car naturalistic driving study. phase 2: Results of the 100-car field experiment. Technical report, United States. Department of Transportation. National Highway Traffic Safety, 2006.

[3] Ali Ghasemzadeh, Britton E Hammit, Mohamed M Ahmed, and Hesham Eldeeb. Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the shrp2 naturalistic driving study & roadway information database. *Safety Science*, 119:21–28, 2019.

[4] Rob Eenink, Yvonne Barnard, Martin Baumann, Xavier Augros, and Fabian Utesch. Udrive: the european naturalistic driving study. In *Proceedings of Transport Research Arena*. IFSTTAR, 2014.

[5] Mohamed M Ahmed and Ali Ghasemzadeh. The impacts of heavy rain on speed and headway behaviors: an investigation using the shrp2 naturalistic driving study data. *Transportation research part C: emerging technologies*, 91:371–384, 2018.

[6] Yvonne Barnard, Fabian Utesch, Nicole van Nes, Rob Eenink, and Martin Baumann. The study design of udrive: the naturalistic driving study across europe for cars, trucks and scooters. *European Transport Research Review*, 8(2):14, 2016.

[7] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2015.

[8] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27(5):1413–1432, 2017.