|  |  |
| --- | --- |
|  |  |

Is pleased to announce

**Modeling Truck Safety Critical Events:**

**Efficient Bayesian Hierarchical Statistical and Reliability Models**

Public and Oral Defense of the Dissertation

**Background:** Large trucks are the primary concern of traffic safety since they are associated with more catastrophic accidents. Since real crashes are very rare, a growing number of studies try to gain insights from naturalistic driving study (NDS) data sets, in which real-world driving data are continuously collected by unobtrusive devices and safety-critical events (SCEs) are used as surrogate measures of real crashes. However, little is known about the value of commercial truck driver NDS datasets and the data analysis framework for data cleaning, aggregation, statistical modeling.

**Objectives:** 1)examine the association between truck crashes and SCEs using Bayesian negative binomial regression models; 2) aggregate NDS data into analyzable units and build scalable hierarchical models to identify potential risk factors for SCEs; 3) identify the pattern of SCEs in shifts using Bayesian recurrent event models.

**Methods:** 1,494,678,173 pings, 34,884 crashes, and 450,75 SCEs were collected from 31,828 commercial truck drivers. We applied Bayesian negative binomial models to examine the association between three outcomes (crashes, injuries, and fatalities) and the four SCEs at the level of drivers. A sample of 496 regional drivers were selected and their 1.3 million pings were aggregated into shifts, trips, and 30-minute intervals. Hierarchical logistic and negative binomial regressions were applied to explain the risk factors for SCEs. Bayesian power law process (PLP) and an innovative jump power law process (JPLP) were proposed to analyze the pattern of SCEs. Large simulations were conducted to assess the performance and validity of these two recurrent-event models.

**Results:** A unit increase in the number of any type of SCEs per 10,000 miles was associated with 8.4% (95% credible interval (CI): 8.0-8.8%) increase in crashes per mile and 8.7% (95% CI: 4.8-13.6%) increase in the number of injuries per mile. Hierarchical logistic and negative binomial models suggested no significant relationship between cumulative driving time and the rate of SCEs. The hierarchical models substantially improved the c-statistics compared to non-hierarchical models. Simulation results show the biases are converging to zero when the number of drivers increases, and the speed of converging is consistent with the central limit theorem. Real data analyses did not identify significant patterns of SCE distribution during shifts.

**Implications for public health:** SCEs can be used as proxies for crashes among truck drivers. The ping 🡪 shift 🡪 trip 🡪 30-minute interval workflow provides an efficient NDS data aggregation framework. Driver-level characteristics are important in explaining SCEs, and hierarchical models that naturally account for this nested nature is recommended for analyzing NDS data sets.

**Miao Cai, PhD (c)**

PhD PHS, Biostatistics

Mentor: Steven E. Rigdon, PhD

Committee Members: Hong Xian, PhD

Fadel M. Megahed, PhD

**Monday, July 20, 2020**

10:00 AM-12:00 PM

Zoom: <https://slu.zoom.us/j/92852200470>