Analysis of differential crash and near-crash involvement based on naturalistic driving data

Wenyan Huang[1]
Johan Engström[1]
Andrew Miller[1]
Felix Dreger[2]
Susan Soccolich[1]
Joost de Winter[2]
Sahar Ghanipoor Machiani[3]

1. Virginia Tech Transportation Institute, U.S.A
2. Delft University, The Netherlands
3. San Diego State University, U.S.A

The goal of this project was to analyze the degree of differential crash/near crash (CNC) involvement in SHRP2.

Differential crash involvement

Some drivers are more likely to become involved in crashes than others.

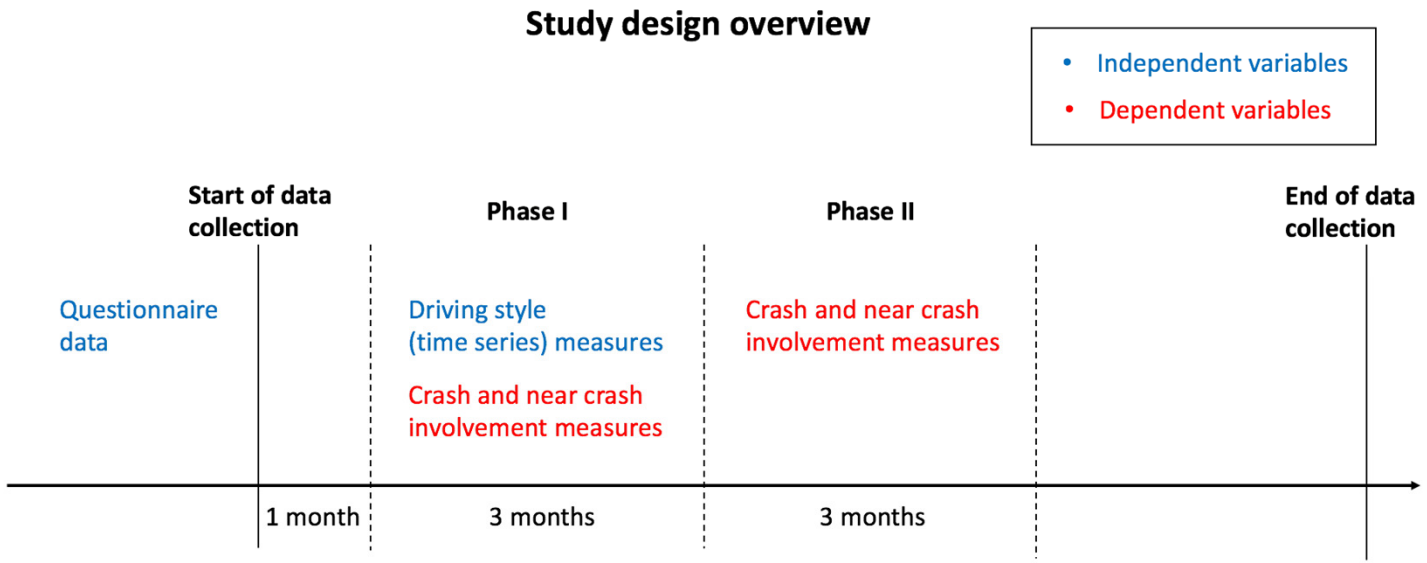
Enduring personal factors

Such differences have been shown to be related to *enduring personal factors* such as demographics (e.g., age, gender), (lack of) driving skills, acquired driving habits (driving style), health issues, and personality-related factors.

Temporary or situational factors

However, crashes may also be associated with more temporary driver or situational factors.

1809 drivers were selected for the analysis after criteria screening.



Drivers were required to have:

- at least seven months participation;
- more than 1,000 driving miles in both Phase I and Phase II.

Trips were required to have:

- more than 10 seconds of moving time;
- a non-zero driving distance;
- date and time information.

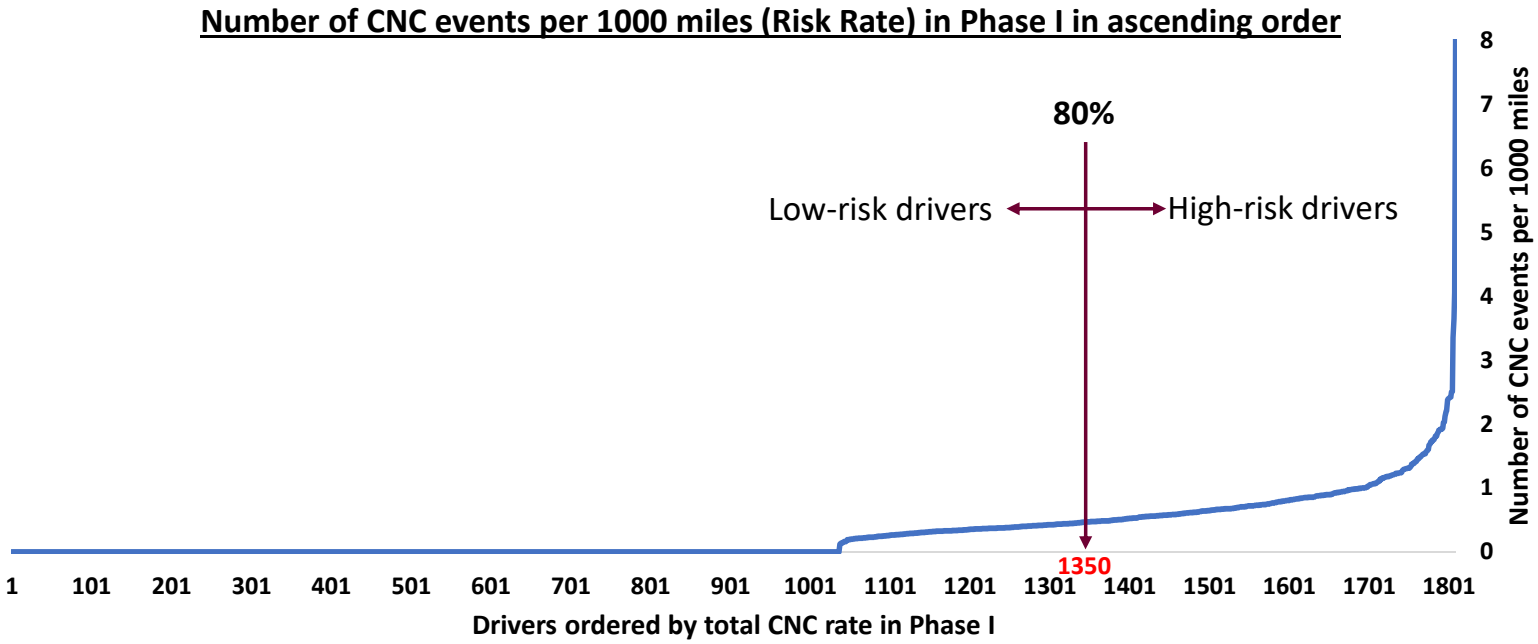
The independent variables in the analysis represented characteristics of individual drivers

- *Demographic:* Age and gender
- *Driving history:* Self-reported violations and crashes in the past two years
- *Personality:* Modified Manchester Driver Behavior Questionnaire (DBQ scale 1 – slips, DBQ-scale 2 – violations, DBQ scale 3 - lapses);
  Sensation Seeking Questionnaire (SSQ total score);
  Risk-perception Behavior Questionnaire (Risk-perception)
- *Driving style:* The rates (number per 1000 miles driving distance) of kinematic events (hard starts, stops, left turns, right turns, left yaw movement, right yaw movement)

Driving style measures in the study period were calculated based on specific kinematic thresholds with the minimum Akaike Information Criterion (AIC) value for the dependent variable.

The dependent variables was binary, high-risk and low-risk drivers.

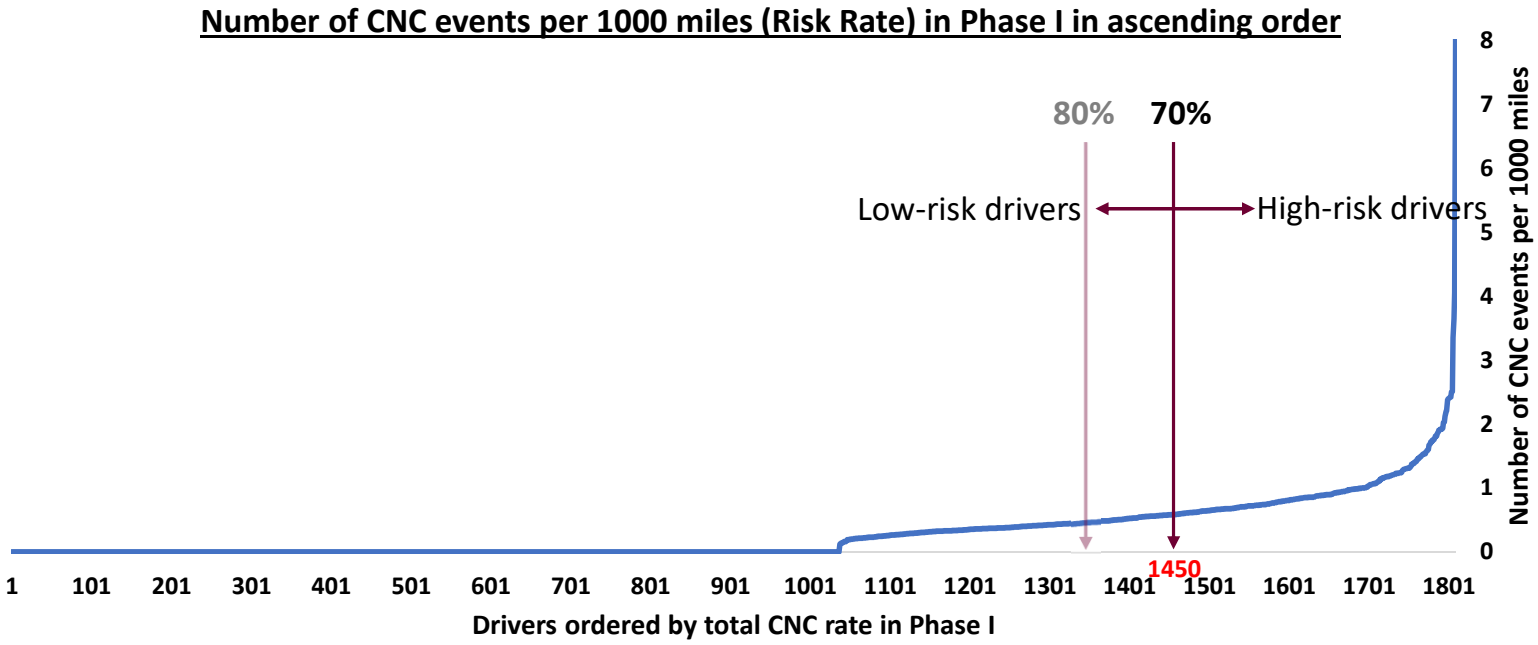
The drivers that accounted for 80% (or 70%, 90%, 95%) of the total CNC rate (total risk) in Phase I or Phase II were classified as ***high-risk drivers*** in the corresponding Phase. The remaining drivers were classified ***low-risk drivers***.



Chi-squared and Wilcoxon rank sum statistical tests were used to examine the difference in each independent variable between high-risk and low-risk drivers.

The dependent variables was binary, high-risk and low-risk drivers.

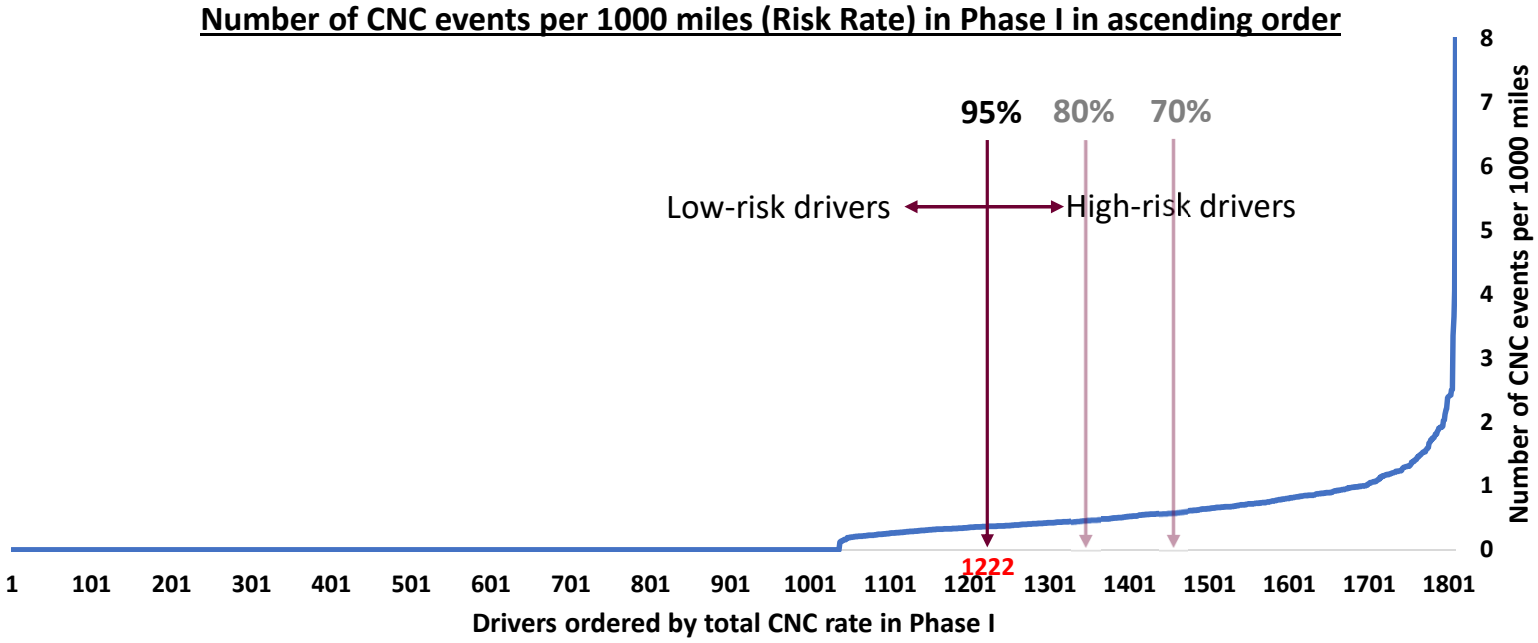
The drivers that accounted for 80% (or 70%, 90%, 95%) of the total CNC rate (total risk) in Phase I or Phase II were classified as ***high-risk drivers*** in the corresponding Phase. The remaining drivers were classified ***low-risk drivers***.



Chi-squared and Wilcoxon rank sum statistical tests were used to examine the difference in each independent variable between high-risk and low-risk drivers.

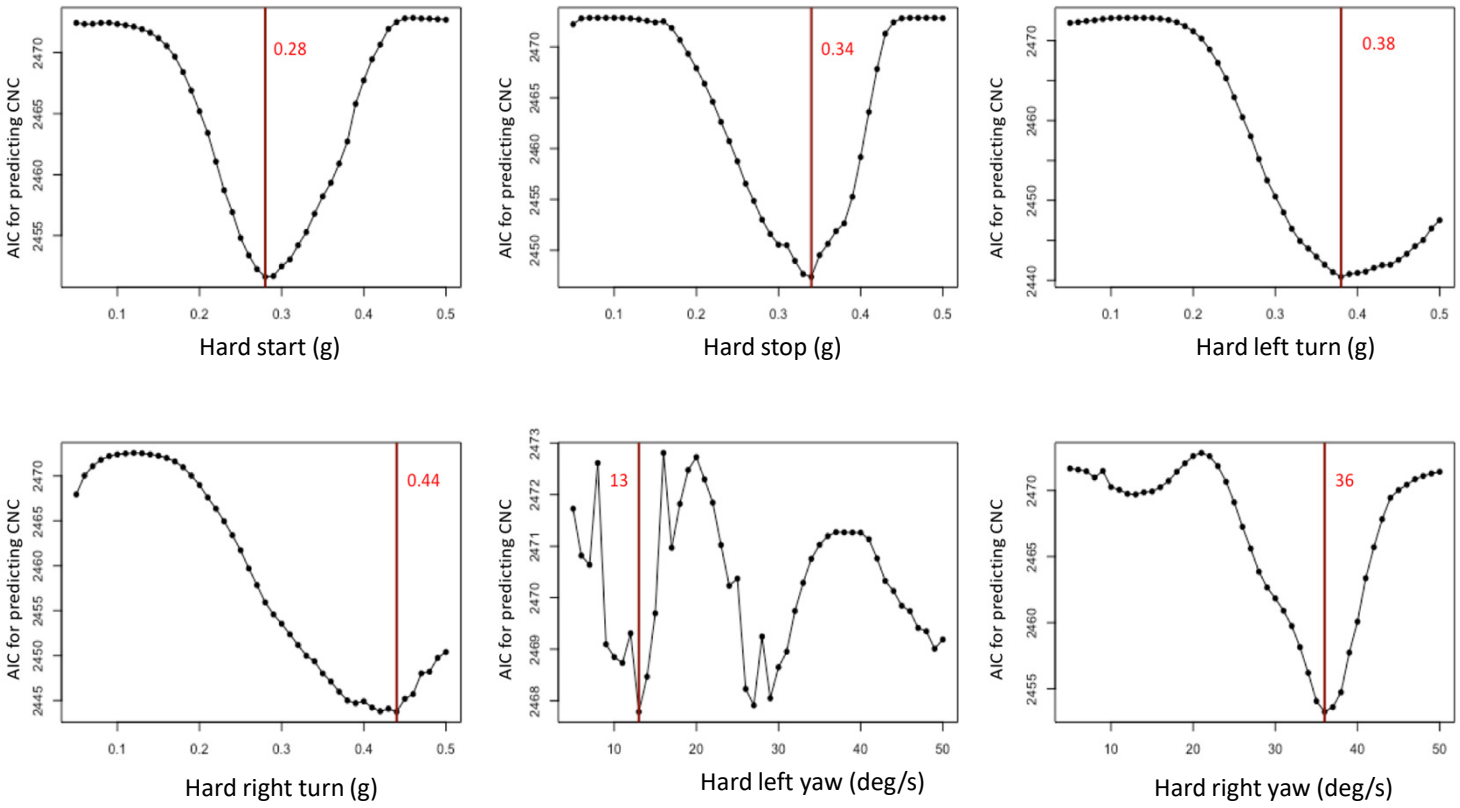
The dependent variables was binary, high-risk and low-risk drivers.

The drivers that accounted for 80% (or 70%, 90%, 95%) of the total CNC rate (total risk) in Phase I or Phase II were classified as ***high-risk drivers*** in the corresponding Phase. The remaining drivers were classified ***low-risk drivers***.



Chi-squared and Wilcoxon rank sum statistical tests were used to examine the difference in each independent variable between high-risk and low-risk drivers.

Specific g-force thresholds results for the driving style measures were selected with minimum AIC value.



Note. Red lines indicate the selected gravitational force level with the minimum AIC value for each driving style measure

Age, DBQ2 violations, and SSQ total score have the highest effect size on differential CNC involvements.

Variable	Type of data	Tests of group differences	Sample size	Effect size	Effect description	
Age	Categorical	Chi-squared test	1,799	**0.13***	⇓	High-risk drivers have a higher proportion of drivers in the younger age groups and lower proportion of drivers in the middle and senior age groups.
Gender			1,809	0.005	⇔	Non-significant effect
Self-reported violations			1,807	0.06*	⇑	High-risk drivers have a higher proportion of drivers with at least one self-reported violation in the past three years.
Self-reported crashes			1,809	0.003	⇔	Non-significant effect
DBQ 1 - slips	Continuous	Wilcoxon rank sum test	1,800	0.08*	⇑	High-risk drivers have a significantly higher average DBQ1 score than low-risk drivers
DBQ 2 - violations			1,800	**0.13***	⇑	High-risk drivers have a significantly higher average DBQ2 score than low-risk drivers.
DBQ 3 - lapses			1,800	0.04	⇔	Non-significant effect
SSQ total score			1,801	**0.13***	⇑	High-risk drivers have a significantly higher average total score than low-risk drivers.
Risk-perception			1,781	0.08*	⇓	High-risk drivers have a lower average risk-perception score than low-risk drivers.

All six driving style measures have significant positive effects on differential CNC involvements.

Variable	Type of data	Tests of group differences	Sample size	Effect size	Effect description	
Hard start rate	Continuous	Wilcoxon rank sum test	1,809	**0.26***	⇑	High-risk drivers have a significantly higher average hard start rate than low-risk drivers.
Hard stops rate			1,809	**0.35***	⇑	High-risk drivers have a significantly higher average hard stop rate than low-risk drivers.
Hard left turns rate			1,809	**0.25***	⇑	High-risk drivers have a significantly higher average hard left turn rate than low-risk drivers.
Hard right turns rate			1,809	**0.22***	⇑	High-risk drivers have a significantly higher average hard right turn rate than low-risk drivers.
Hard left yaws rate			1,809	**0.12***	⇑	High-risk drivers have a significantly higher average hard left yaw rate than low-risk drivers.
Hard right yaws rate			1,809	**0.18***	⇑	High-risk drivers have a significantly higher average hard right yaw rate than low-risk drivers.

Note: <0.1 = negligible effect size; <0.3 = small effect size; < 0.4 medium effect size.

Differential crash involvement is partly related to individual differences.

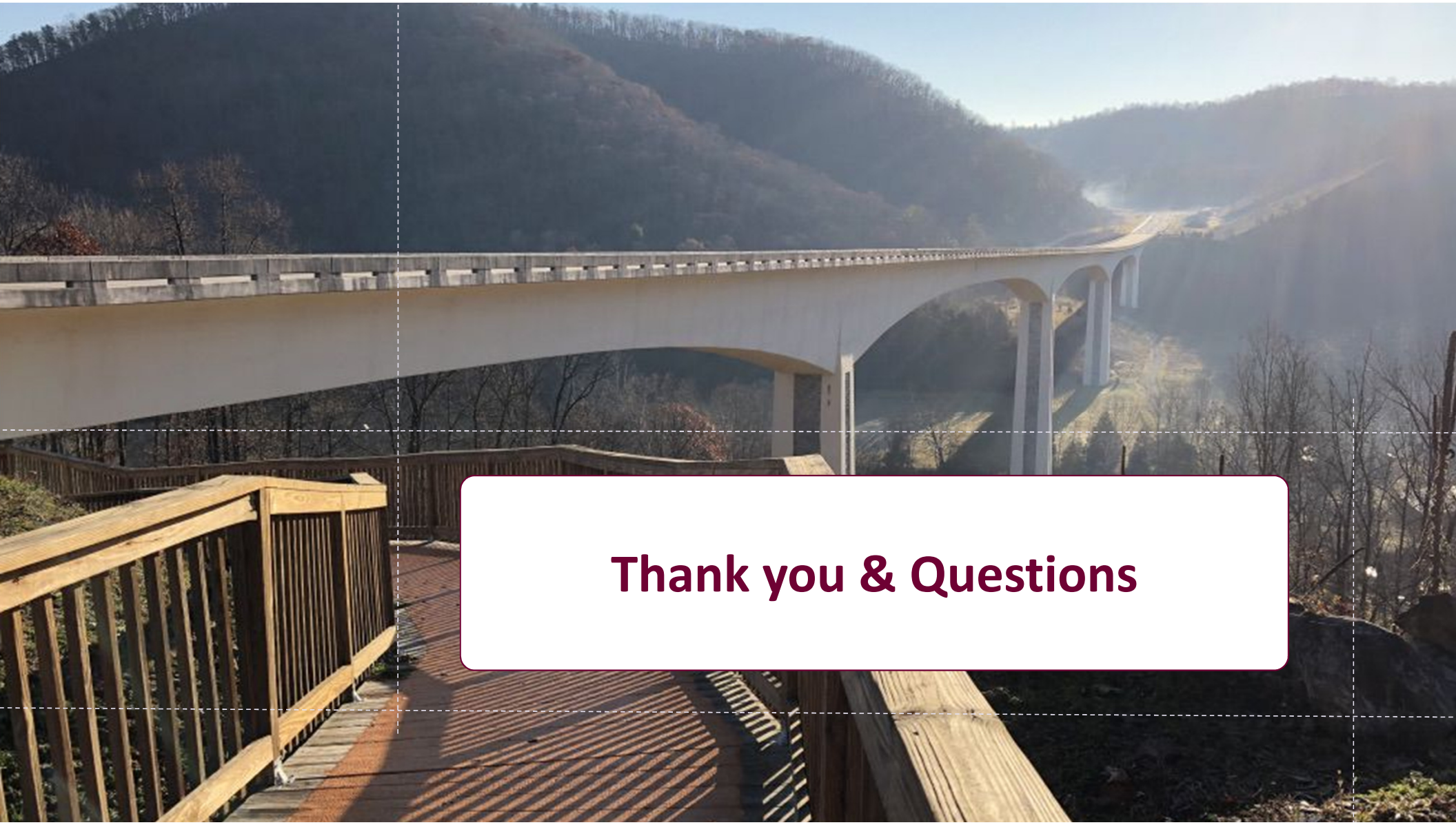
- High-risk drivers (accounting for 80% of the CNCs):
  - are overrepresented in young age groups;
  - have a higher self-reported tendency to commit errors (slips and violations);
  - have a higher self-reported engagement in sensation-seeking behavior;
  - are more likely to have had at least one self-reported violation during the past three years;
  - have a higher rates of kinematic events (including hard starts, stops, left turns, right turns, left yaw movements, right yaw movements).
- These results are significant (p-value less than 0.05), but most of their effect sizes are small (less than 0.3).

CNC involvement is somewhat persistent over time for individual drivers and hence, to some extent, predictable based on enduring personal factors.

Criterion for high risk driver classification (percentage of the total CNC rate accounted for by high-risk drivers)	**Proportions of high/low risk drivers in Phase I**	**Proportions of high/low risk drivers in Phase II**		**Relative Risk**
		Low-risk drivers	**High-risk drivers**	
80%	Low-risk drivers (74.6%)	1107 (61.2%)	243 (13.4%)	**2.23**
	High-risk drivers (25.4%)	275 (15.2%)	184 (10.2%)	(1.90, 2.61)

CNC involvement is somewhat persistent over time for individual drivers and hence, to some extent, predictable based on enduring personal factors.

Criterion for high risk driver classification (percentage of the total CNC rate accounted for by high-risk drivers)	**Proportions of high/low risk drivers in Phase I**	**Proportions of high/low risk drivers in Phase II**		**Relative Risk**
		Low-risk drivers	**High-risk drivers**	
80%	Low-risk drivers (74.6%)	1107 (61.2%)	243 (13.4%)	2.23
	High-risk drivers (25.4%)	**275 (15.2%)**	184 (10.2%)	(1.90, 2.61)


Thank you & Questions

Appendix

The finding that 25.4% of the drivers accounted for 80% of total CNC rate in Phase I provides clear evidence for differential crash involvement.

Criterion for high risk driver classification (percentage of the total CNC rate accounted for by high-risk drivers)	**Proportions of high/low risk drivers in Phase I**	**Proportions of high/low risk drivers in Phase II**		**Relative Risk**
		Low-risk drivers	**High-risk drivers**	
70%	Low-risk drivers (80.2%)	1246 (68.9%)	204 (11.3%)	2.51
	High-risk drivers (19.8%)	232 (12.8%)	127 (7.0%)	(2.08, 3.04)
80%	**Low-risk drivers (74.6%)**	**1107 (61.2%)**	**243 (13.4%)**	**2.23**
	High-risk drivers (25.4%)	**275 (15.2%)**	**184 (10.2%)**	**(1.90, 2.61)**
90%	Low-risk drivers (67.6%)	933 (51.6%)	289 (16.0%)	1.89
	High-risk drivers (32.4%)	325 (18.0%)	262 (14.5%)	(1.65, 2.16)
95%	Low-risk drivers (63.3%)	836 (46.2%)	309 (17.1%)	1.78
	High-risk drivers (36.7%)	345 (19.1%)	319 (17.6%)	(1.57, 2.01)