Analysis and Classification of Crashes and Near-Crashes involving 16-19 year old drivers

Yohan Chang, Nipjyoti Bharadwaj, Praveen Edara, Carlos Sun

University of Missouri - Columbia

7th International Symposium on Naturalistic Driving Research

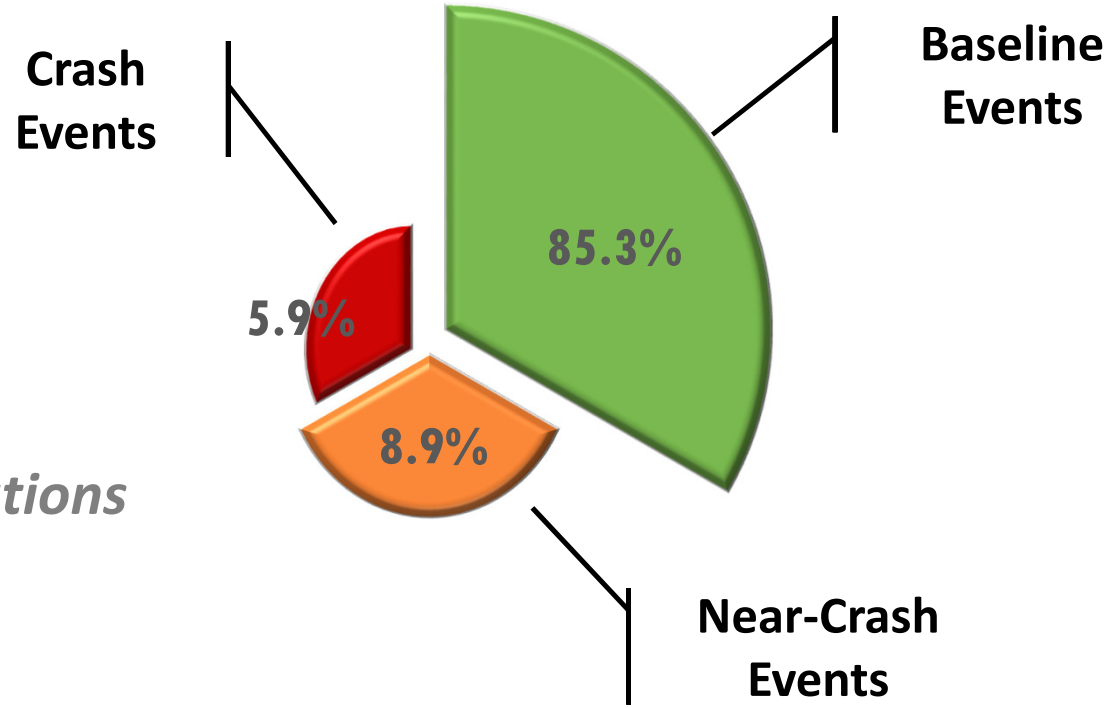
OUTLINE

- **Studying younger driver safety**
- **NDS - Computing the Odds**
- **NDS – Classifying events**
- **Findings**
- **Key takeaways**

Studying Younger Driver Safety

- ✓ 19 young lives lost each day in crashes
- ✓ Understanding younger driver behavior
  - ✓ Two previous NDS studies used 100 car study data
  - ✓ SHRP2 NDS data is a larger sample and more variability
- ✓ Two objectives
  1) Quantify effect of causal factors using odds ratios
  2) Predict events using pre-incident variables

SHRP 2 NDS Data on 16-19 years drivers



- ✓ *6,048 total observations*

Computing the odds

- **Logistic regression** to estimate the younger drivers' risk of crashes and near-crashes (CNC)
- A dichotomous variable was created by combining CNC as one response and baseline as the other.
  - **Younger driver group** vs. **all-age groups**
  - **Explanatory** variables –
    - duration of secondary task;
    - driving behavior;
    - maneuver judgement;
    - traffic density;
    - intersection influence; and
    - miles travelled last year

Odds Ratios – Younger Driver and All-ages

Variable	Odds ratio (95% CI in parenthesis)	
	Model 1: Younger drivers	**Model 2: All-ages**
Duration of secondary task: > 6 secs vs. 0-6 secs	4.84 (3.86-6.08)	4.75 (4.28-5.28)
Behavior : Violation vs. None	2.62 (1.61-4.26)	2.29 (1.87-2.79)
Behavior : Mistake vs. None	6.19 (4.34-8.85)	5.97 (5.18-6.89)
Behavior : Inattention vs. None	**28.70 (21.25-38.76)**	21.10 (18.59-23.97)
Behavior : Inexperience vs. None	30.03 (6.34-142.23)	32.16 (17.45-61.74)
Maneuver : Unsafe vs. Safe	6.45 (4.01-10.36)	5.93 (4.88-7.23)

Variable	Odds ratio (95% CI in parenthesis)	
	Model 1: Younger drivers	**Model 2: All-ages**
Traffic density : LOS* B vs. LOS A	1.57 (1.26-1.96)	1.75 (1.59-1.92)
Traffic density : LOS C vs. LOS A	3.76 (2.59-5.47)	4.06 (3.50-4.70)
Traffic density: LOS D vs. LOS A	4.05 (1.83-8.92)	5.43 (4.27-6.87)
Traffic density : LOS E vs. LOS A	1.83 (0.48-6.93)	2.95 (1.99-4.28)
Traffic density : LOS F vs. LOS A	2.54 (0.49-13.18)	1.95 (0.90-4.01)
Miles travelled last year : 5000 vs. >25000 miles	**2.01 (1.12-3.63)**	
Miles travelled last year : 5000-15000 vs. >25000 miles	1.78 (0.99-3.21)	
Miles travelled last year : 15000-25000 vs. >25000 miles	1.40 (0.77-2.54)	

*LOS – Level of Service (A-best, F-worst traffic conditions)

Predicting events using machine learning

✓ ***Three algorithms*** *to classify and predict safety critical events*

- Random forest (**RF**)
  - Supervised learning + Tabu-search
- Deep Neural Network (**DNN**)
  - Semi-supervised learning + Genetic algorithm (GA)
- t-Distributed Stochastic Neighbor Embedding (**t-SNE**)
  - Unsupervised learning + Tabu-search and K-nearest-neighbors (KNN) algorithm

Prediction scenarios

- ✓ **Ordinary encoding** for RF and DNN, **ASCII encoding** for t-SNE model
- ✓ **Two scenarios for classification algorithm**

Scenario	Classification	Input Variables
Case-I	Crash/Near-Crash, vs. Baseline events	10 driver characteristics + 31 pre-incident variables
Case-II	Crash, vs. Near-Crash events	10 driver characteristics + 62 pre-incident variables

Prediction results

Results for test data (random 25% of entire sample)

	Naïve	**RF**	**t-SNE**	**DNN**
Case-I: **Crash/Near-Crash vs Baseline**	85.25%	94.71%	87.37%	86.23%
Case-II: **Crash vs Near-Crash**	60.64%	87.00%	85.85%	60.64%

Prediction results

Variable importance from the RF model

Case-I	% importance
Driver behavior	38.8%
Secondary task duration	23.3%

Case-II	% importance
Vehicle 2 configuration	15%
Event nature	7.0%
Precipitating event	6.3%

Key takeaways

- ✓ Inattention and inexperience cause high level of crash risk
  - ✓ In younger drivers the effect of inattention is more pronounced
- ✓ Traffic conditions approaching LOS D increased crash risk
- ✓ More driving experience decreases risk in younger drivers
- ✓ Random Forest method accurately predicted NDS events
  - ✓ Run times low to support in-vehicle driver advisory systems

Thank you

Acknowledgments:

- Charles Fay, FHWA, Project Manager for SHRP 2 BAA grant that supported this study
- VTTI team for providing the NDS data used in the study

For further information, contact:

Praveen Edara, edarap@missouri.edu, 573-882-1900