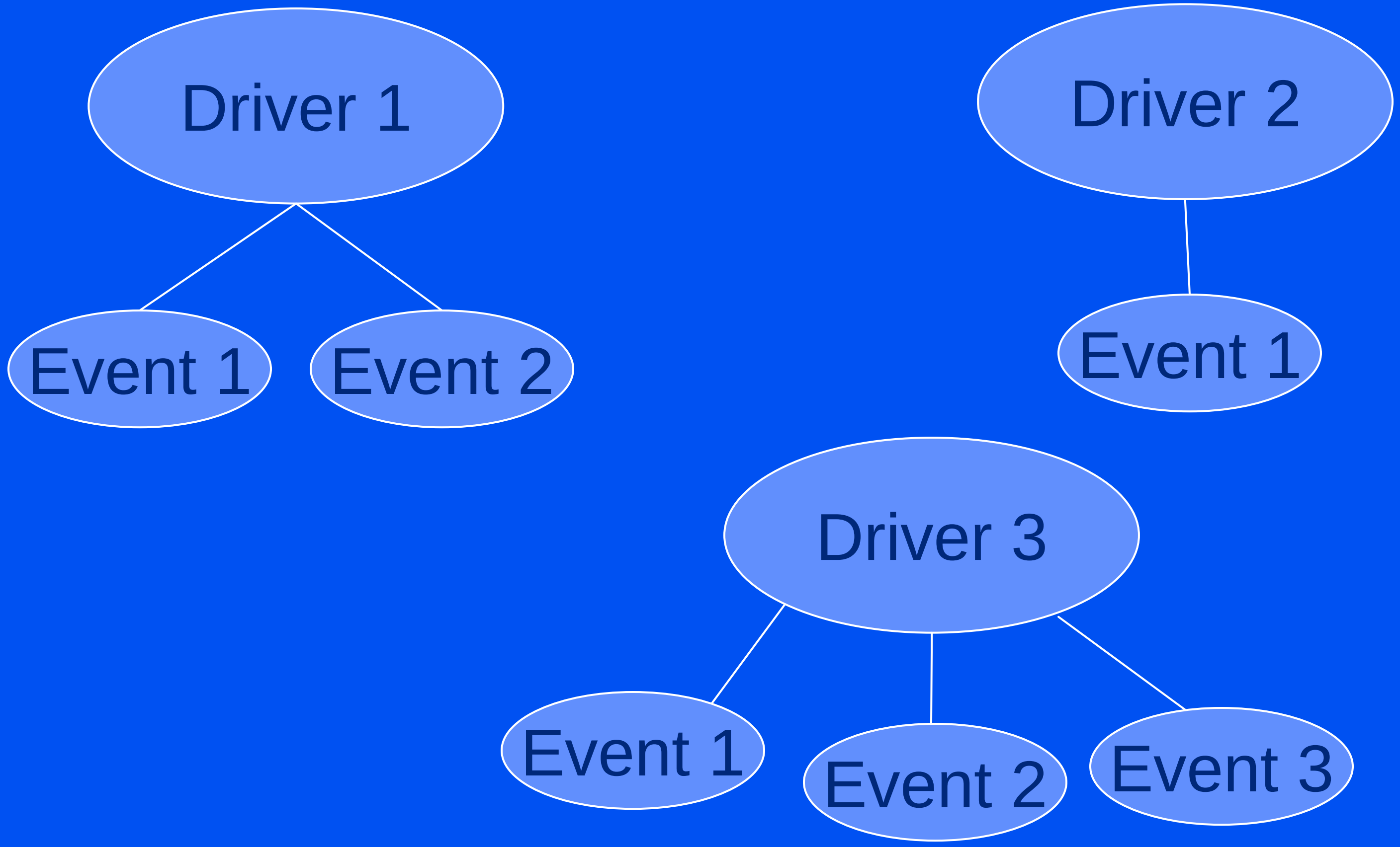
SHRP 2 S-01b
Analysis of Existing Data:
"Prospective Views on Methodological Paradigms"

Paul P. Jovanis, Venky Shankar, Jonathan Aguero, Kun-Feng Wu, Adam Greenstein
Penn State University
Civil and Environmental Engineering
Larson Transportation Institute

Key Findings

- Identify and verify crash surrogate events
- Identification of high risk drivers
- Prediction differences: Bayesian vs. Frequentist Models

Bayesian Model Structure



Event-Based Model

- $Y_{ij} \sim$ *Bernoulli* (p_{ij})
- $Y_{ij} \longrightarrow$ *Event/Context i for Driver j*

[Crash/Near Crash] $\longrightarrow 1$; *Incident* $\longrightarrow 0$

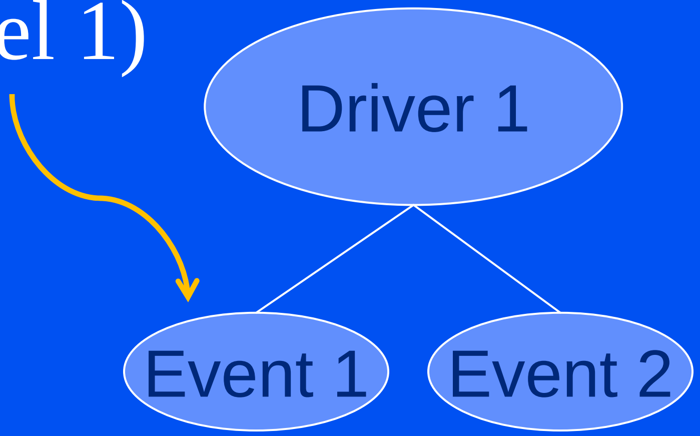
- $X_i \longrightarrow$ *Event/Context Characteristics*

$Z_j \longrightarrow$ *Driver Characteristics*

- $Logit\ (p_{ij}) = \alpha + \boldsymbol{\beta}_i \boldsymbol{X}_i + \gamma_j \boldsymbol{Z}_j$
- *Note: Only drivers with C/NC/I are represented in the model (~ 60% of drivers)*

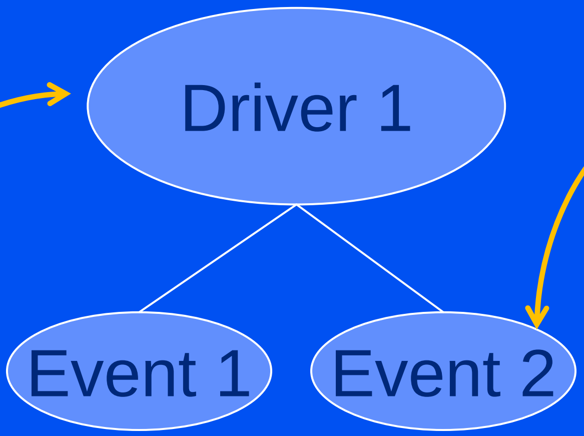
Surrogate Analysis – VTTI Data

- Run-off-Road events: 17 crashes; 30 near-crashes and 150 critical incidents
- Event predictors variables (Level 1)
  - *Precipitating Event:* Lost Control; Subject over lane line/road edge
  - *Driver Impairment*: drowsy, sleepy, asleep, & fatigue
  - *Distraction*: Wireless device; Vehicle-related; Passenger-related; Talking/singing/daydreaming; Internal distraction; Dining; Other



Surrogate Analysis – VTTI Data

- Context (Level 1): alignment: curve; lighting; pavement condition; traffic density



- Driver attributes (Level 2): DULA driving index; Years driving; Life Stress Index

C/NC vs. Critical Incident

VARIABLE	PARAM.	S.D.	SIG.	ODDS RATIO
Intercept	-1.76	1.89		-
Precipitating Event1: Lost of Control	1.51	1.13	.	9.41
Precipitating Event2: Subject over lane line/road edge	2.82	1.11	*	35.35
Driver Impairment1: drowsy, sleepy, asleep, & fatigue	1.29	0.59	*	4.26
Distraction1: Wireless device	0.25	0.78		1.73
Distraction2: Vehicle related	1.98	0.96	*	11.29
Distraction3: Passenger related	1.53	0.85	*	6.67
Distraction4: Talking/singing/daydreaming	1.67	1.13	.	9.92
Distraction5: Internal distraction	3.17	1.01	*	41.32
Distraction6: Dining	1.05	1.46		8.21
Distraction7: Other	1.14	1.06		5.49

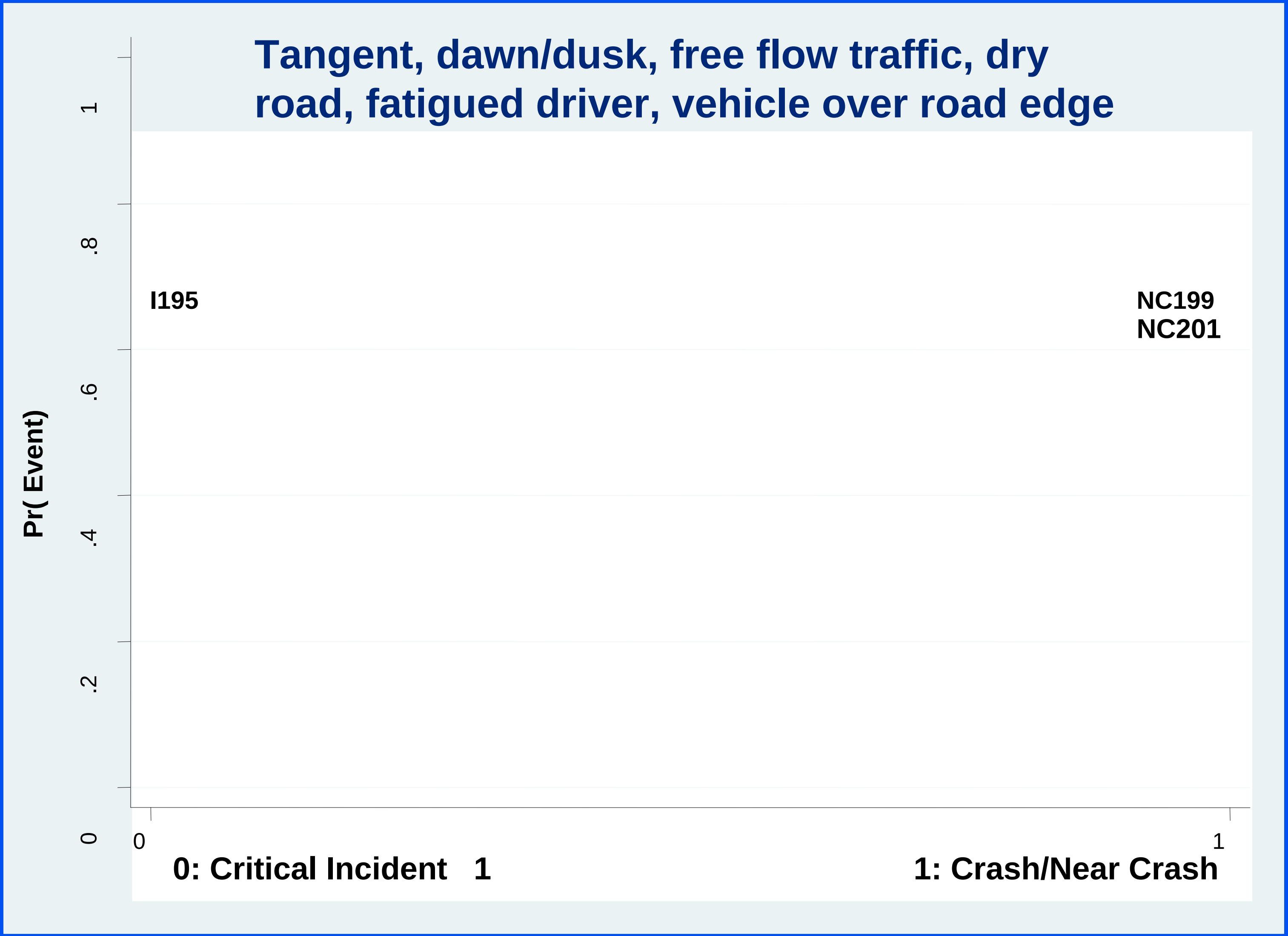
C/NC vs. Critical Incident

VARIABLE	PARAM.	S.E.	SIGNIF.	ODDS RATIO
Alignment1: Curve	1.10	0.51	*	3.43
Lighting1: Dawn/Dusk	2.42	0.79	*	15.40
Surface condition1: Wet/Icy/Snowy	0.82	0.64	•	2.80
Traffic density1: not free flow	-2.40	0.70	*	0.11
DULA Aggressive Driving Index	-0.14	0.05	*	0.87
DULA Negative Emotions Index	-0.12	0.08	•	0.89
DULA Reckless Driving Index	0.07	0.10		1.08
Driver Experience	-0.07	0.02	*	0.94
Life Stress Index	0.01	0.00	*	1.01

Predicted Incidents and Crashes

Predicted Incidents and Crashes

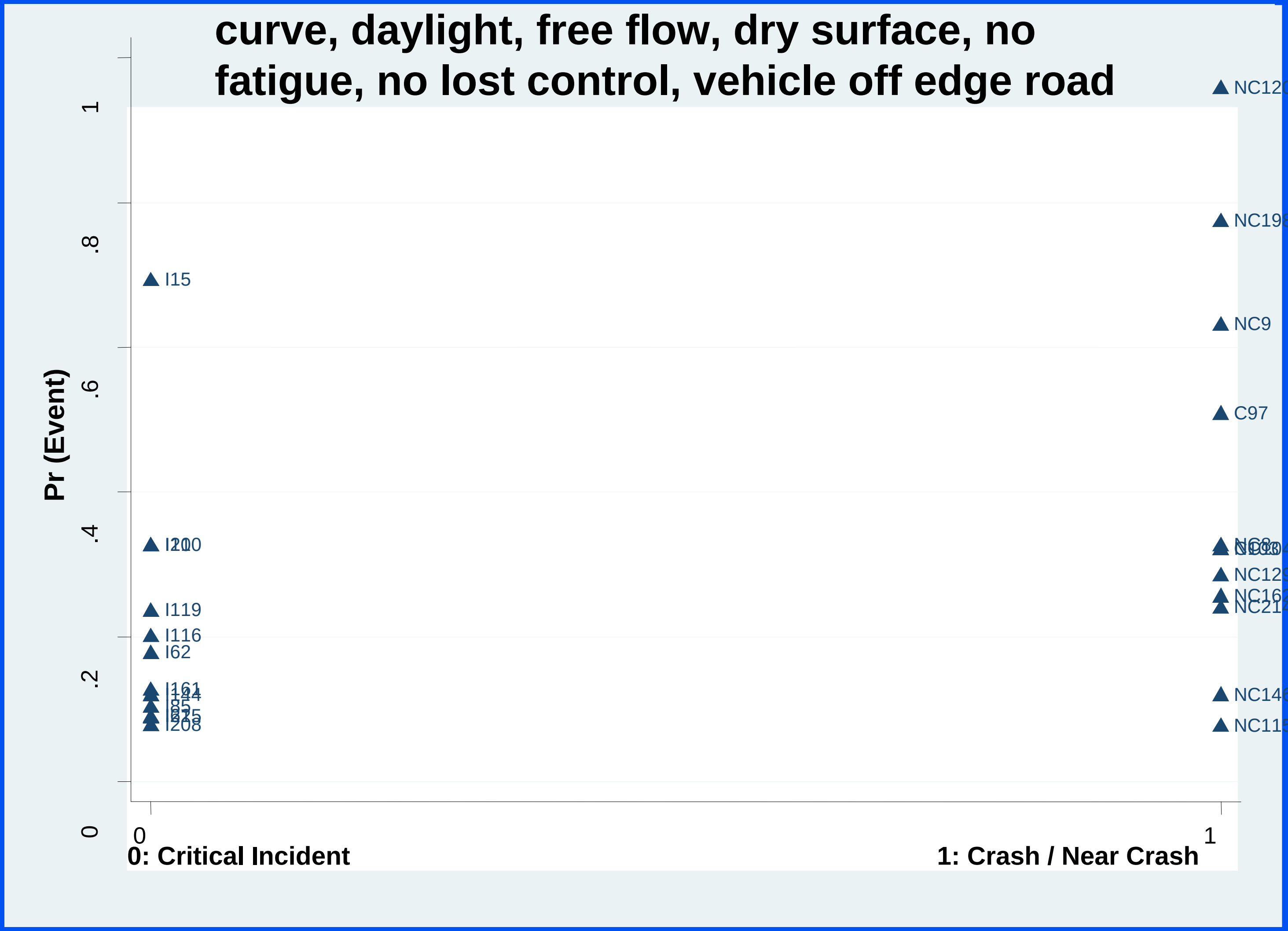
Model Prediction: Actual Context



Verification of Event Similarity

- **I195**: The subject fell asleep behind the wheel and drifted towards the right edge of the road. He suddenly woke himself up and jerked the wheel to the left to get back in his lane.
- **NC199:** Subject driver falls asleep while driving and the vehicle runs off the road on the right.
- **NC201:** Subject driver falls asleep while driving and the vehicle runs off the road on the right.

Using Model Output

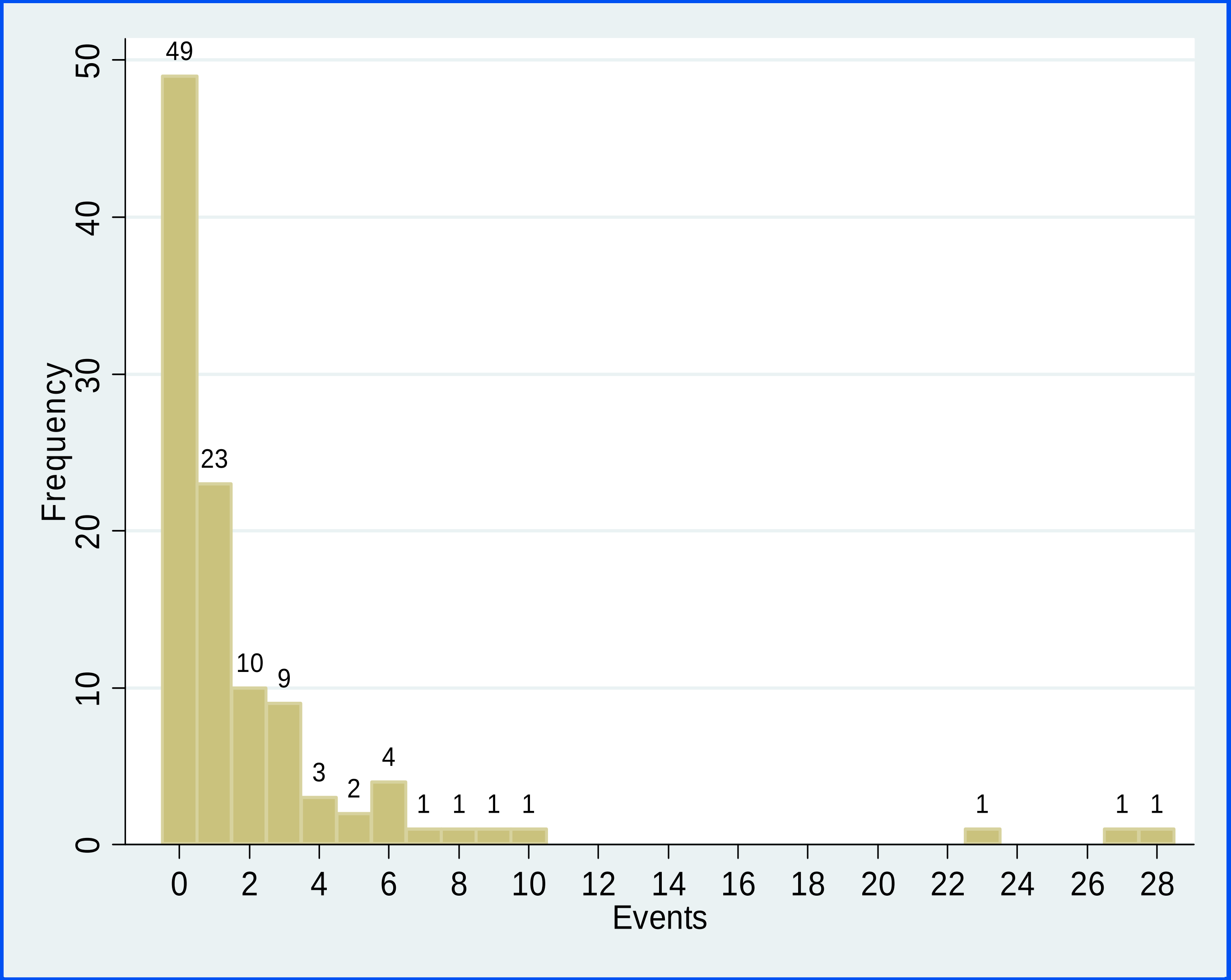


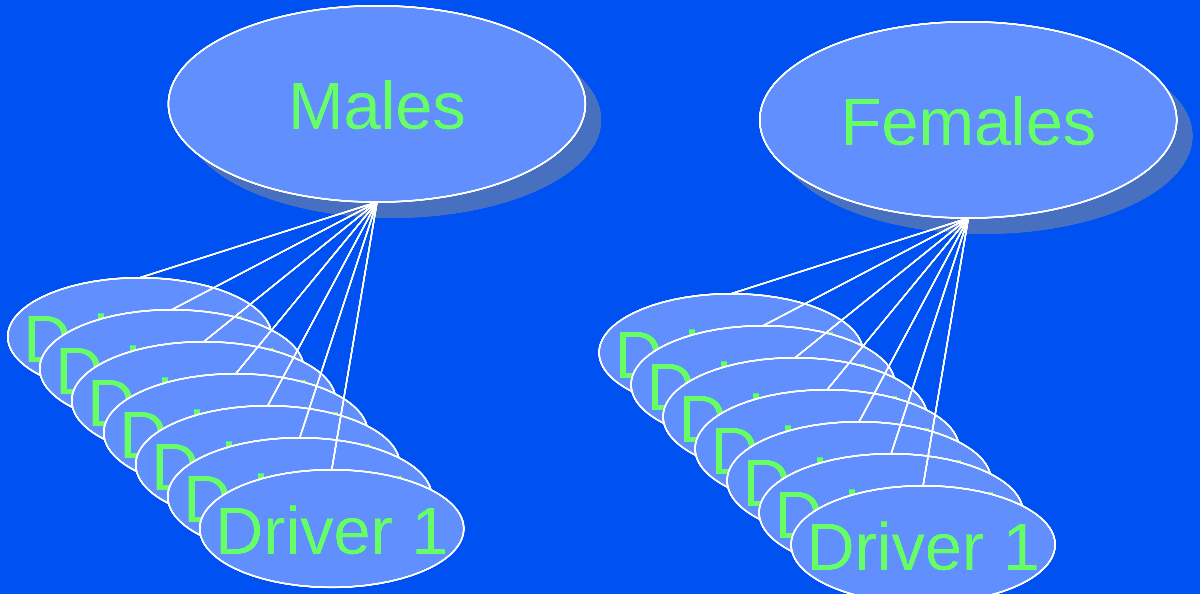
Bottom Line

- Quantitative method to identify promising surrogate events using event-based models
- Alternative specifications possible
  - Crashes vs. critical incidents
  - Additional contexts
- Event, context and driver attributes important

Identification of Outliers or Drivers With Promise (DWiPs)

Number of Events per Driver




Males
Females
Driver 1
Driver 1

Driver-Based Hierarchical Model

Y_{ij} ~ Poisson (λ_{ij}) number of events, driver i, gender j

$$Log\ (\lambda\ ) \sim \alpha_j + \boldsymbol{\beta_j}\ \boldsymbol{X_{ij}} + v_{ij}$$

$\boldsymbol{X_{ij}}$ $\longrightarrow$ *Level 1 covariates (driver attributes) such as mileage, education level, Dula score, driving experience, etc.*

V_{ij} $\longrightarrow$ *driver-based random effects*

$\longrightarrow$

Driver-Based Hierarchical Model

Y_{ij} ~ *Poisson* (λ_{ij}) *number of events, driver i, gender j*

$Log\ (\lambda\ ) \sim \alpha_j + \boldsymbol{\beta_j X_{ij}} + v_{ij}$

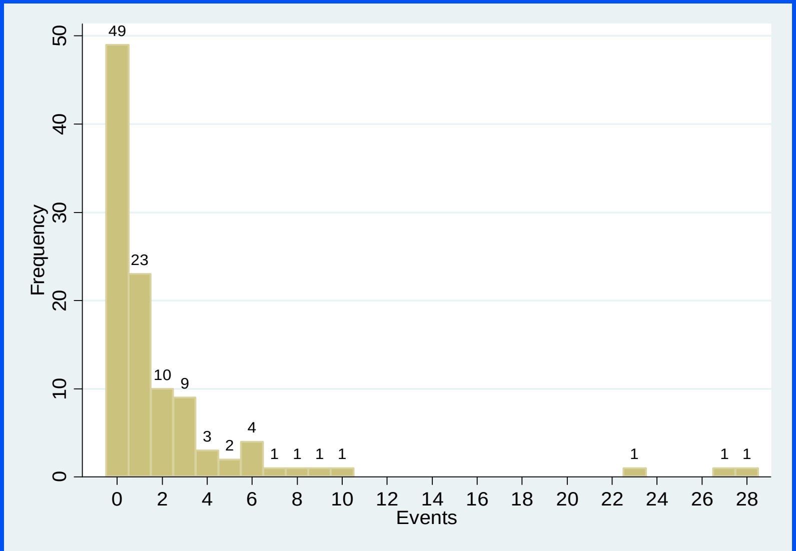
$\boldsymbol{X_{ij}}$ ⟶ *Level 1 covariates (driver attributes) such as mileage, education level, Dula score, driving experience, etc.*

V_{ij} ⟶ ***driver-based random effects***

⟶

Identify High Risk Drivers

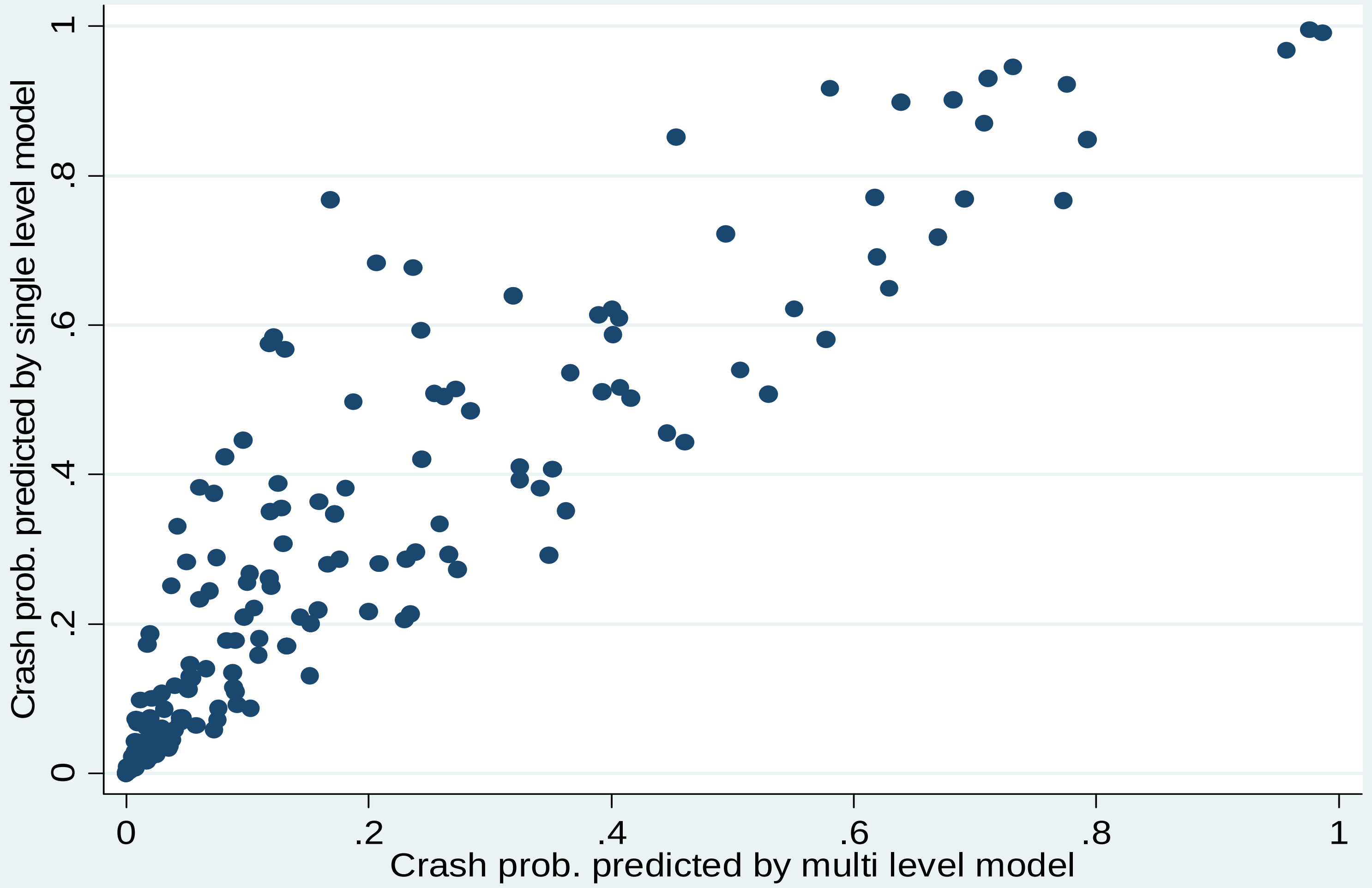
Five drivers have V_{ij} different from zero



	mean	S.D.	2.50%	5%	10%	90%	95%	97.50%
v[2]	1.78	0.92	0.03	0.31	0.63	2.96	3.33	3.66
v[4]	-1.75	0.82	-3.45	-3.15	-2.81	-0.73	-0.45	-0.21
v[15]	1.80	0.60	0.64	0.84	1.05	2.57	2.80	3.01
v[16]	2.01	0.62	0.83	1.02	1.23	2.80	3.05	3.26
v[55]	2.59	0.61	1.42	1.61	1.83	3.36	3.60	3.81

Differences Between Bayesian and Frequentist Model Estimates

Bayesian vs. Frequentist



Planned Next Steps

- Continued modeling of surrogate events
- Differences between Bayesian and frequentist model estimates
- Driver adaptation to Curve Speed Warning (CSW) and Lateral Drift Warning (LDW)
- Explore effect of pre-disposition measures
- Test UMTRI data structure with cohort-based methods.