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

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

Overview of Presentation

- Proposal Theme
- Data in Hand
- Review Progress Concerning Four Research Questions
- Findings to Date
- Planned Next Steps

Theme of our Proposal

- In addition to understanding details of events immediately prior to crash, identify methodological paradigms that can be used to answer questions long of interest to safety researchers

VTTI Data in Hand

Run-Off Road Events

	Received from VTTI	Used in Modeling
Crashes	22	17
Near-Crashes	35	30
Critical incidents	171	150
Total	228	197
Drivers	107	83

Received End February 2008

In Addition: 4412 Baseline Events

Research Question 1

- What is the nature of the relationship between crashes, near-crashes, incidents and pre-event maneuvers, precipitating factors, driver factors, contributing factors and environmental factors?

Research Question 1

- What is the nature of the relationship between crashes, near-crashes, incidents and pre-event maneuvers, precipitating factors, driver factors, contributing factors and environmental factors?

Driver-Based Model

- **Probability of Y_i Events for Driver i**
- ***Y_i ~ Negative Binomial***
- ***Driver attributes:***
  - ***Objective (e.g. years driving; annual mileage)***
  - ***Subjective (pre-disposing) – Dula AD***

Driver-Based Model

Negative Binomial Regression – Probability of Y_i Events for Driver i

Variable	Coef.	SE	Z	P > |Z|
Male Driving Less than 10 years	1.22	0.48	2.57	0.01
Vehicle Age and Female	-0.31	0.14	-2.20	0.03
Male with Bachelor's Degree or more	-1.13	0.51	-2.22	0.03
Number of Previous Violations and Female	-0.55	0.21	-2.62	0.01
Annual Miles Driven (thousands miles)	0.15	0.03	4.67	0.00
Scaled Dula AD and Female	0.52	0.21	2.51	0.01
Constant	-1.01	0.69	-1.45	0.15
Alpha	1.11	0.28		

McFadden $R^2 = 0.130$

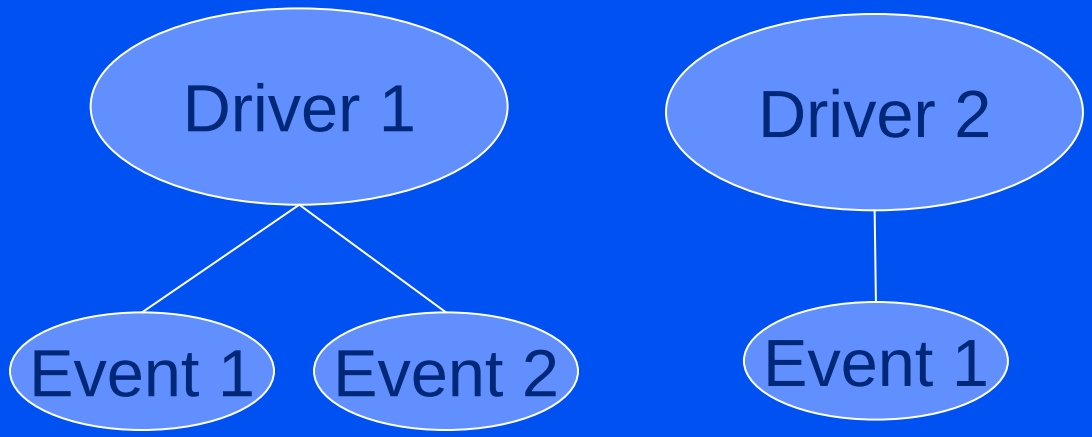
Model Interpretation

- Expected number of events increases with annual mileage driven
- Females with older vehicles and a record of previous violations have a reduced likelihood of an event
- Males with less than 10 years experience have an increased risk of an event; those with BS and above have reduced frequency
- Females with high Aggressive Driving ratings have increased likelihood of events

Research Question 2

- What hierarchical structure (statistically speaking) if any, exists in the manner in which these relationships need to be explored?

Event-Based Model - Hierarchical



Event-Based Model

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

[Crash/Near Crash] —1; *Incident* —0

- $X_i \longrightarrow$ *Event Characteristics*

Z_j —*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 (~ 70% of drivers)*

Event-Based Model

Variable Name	Mean	S.D.	5%	95%	Signif.?
Intercept	-1.302	1.795	-4.35	1.61	
Precipitating Event – Lost control	1.396	1.121	-0.30	3.34	.
Precipitating Event – Subject over lane line/road edge	3.027	1.115	1.37	4.99	*
Distraction 1– Wireless device	-0.237	0.751	-1.49	0.98	
Distraction 2– Vehicle-related	1.402	0.934	-0.15	2.91	.
Distraction3 – Passenger-related	1.154	0.840	-0.22	2.54	.
Distraction4 - Talking/singing/daydreaming	1.197	1.101	-0.65	2.97	
Distraction5 – Internal distraction	2.863	0.979	1.31	4.52	*
Distraction6 - Dining	0.665	1.469	-1.78	3.04	
Distraction7 - Other	0.928	1.049	-0.80	2.64	
Behavior – Improper speed	-0.131	0.644	-1.21	0.91	

Event-Based Model

Variable Name	Mean	S.D.	5%	95%	Signif.?
Surface conditions – wet/snowy icy	0.766	0.633	-0.27	1.82	
Traffic density not free flow	-2.302	0.676	-3.46	-1.25	*
Alignment – curve	1.109	0.504	0.29	1.95	*
Lighting – dawn/dusk	2.504	0.771	1.24	3.78	*
Dula – Aggressive Driving	-0.137	0.054	-0.23	-0.05	*
Dula – Negative Emotion	-0.132	0.076	-0.26	-0.01	.
Dula – Reckless Driving	0.079	0.097	-0.08	0.24	
Weighted Life Stress	0.0050	0.0016	0.00	0.01	*
Years Driving	-0.059	0.020	-0.09	-0.03	*

Event-Based Model Discussion

- Precipitating events are significant: “Loss-of-Control” (marginally) and “Subject-over-lane-line/road-edge”.
- Distraction variables
  - Baseline is no distraction
  - *Distraction1 Wireless device*
  - ***Distraction2: Vehicle related***
  - ***Distraction3: Passenger-related***
  - *Distraction4: Talking/singing/daydreaming*
  - ***Distraction5: Internal distraction***
  - *Distraction6: Dining*
  - *Distraction7: Other*
- Improper speed not significant

Event-Based Model Discussion

- Context:
  - traffic density – higher traffic flow is higher, reduces the likelihood of a run-off-road crash/near crash.
  - Curve and conditions at dawn/dusk increase run-off-road crash/near crash likelihood.
  - Surface condition significant in later models
- Driver :
  - DULA AD Intent to harm – negative so reduces C/NC and increases I - *"I verbally insult drivers who annoy me";*
  - DULA NE – negative also - *Passengers in my car/truck tell me to calm down";*
  - DULA RD - *"I will weave in and out of slower traffic"*
  - LIFE STRESS and YEARS DRIVING are significant in differentiating events
- **HAVE CONTRIBUTING FACTORS FROM ALL 3 COMPONENTS : DRIVER, EVENT AND CONTEXT**

Research Question 3

- What kind of elucidative evidence emerges from the analysis of roadway departure crashes in terms of questions 1 and 2? Is the illustrative hierarchy of relationships generalizable to other non-intersection crash types such as leading vehicle crashes, for example?
- *Seeking methods to gain insight despite data limitations*

Research Question 4

- In terms of elucidative evidence, what types of behavioral correlates emerge? For example, are attitudinal (e.g. “predisposing”) measurements indicative of revealed behavior in terms of headway maintenance and speed reductions?
- *Have begun to explore relationships w.r.t. C/NC/I*

Evolving Research Issues

- Association between C/NC and I outcomes
  - Need for refined, agreed-upon definitions
  - Importance in 2500-car data collection
- Driver “predisposition” measurable and significant
- Clear gender differences (as expected); implications for experimental design of 2500-car
- Continue to explore alternative hierarchical models in respond to research questions
- Incident data seem valuable and add insight concerning crashes and surrogates

Planned Next Steps

- Analyze model residuals
- Assess magnitude of importance of predictors
- Explore additional hierarchical model structures
- Explore use of VTTI baseline data
- Complete UMTRI data sharing agreement; obtain data
- Analyze UMTRI data using full suite of models
- Test UMTRI data structure with cohort-based methods.

Hierarchical Model II

$Y_{ij} \sim Poisson(\lambda_{ij})$

$Y_{ij} \rightarrow$ number of conflicts

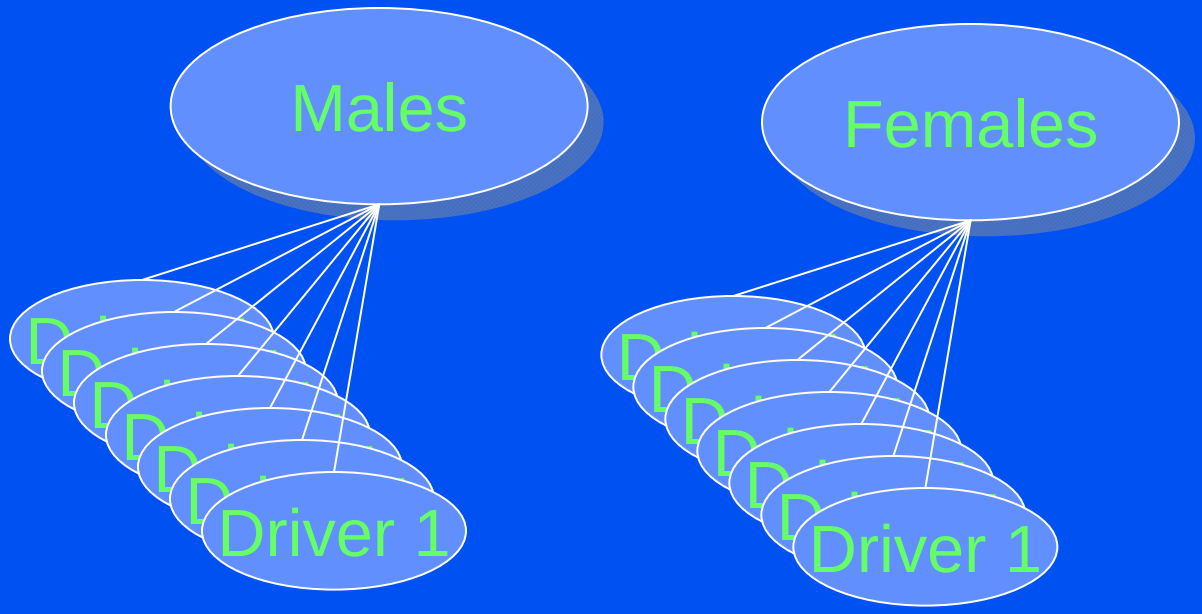
$i \rightarrow$ driver charateristics

$j \rightarrow$ gender

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

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

$u_{ij} \longrightarrow$ *variability between gender groups*


Males
Females
Driver 1
Driver 1

Hierarchical Poisson

	Mean	Standard Deviation	Sig.?
Intercept[F]	-13.71	8.26	*
Intercept[M]	-18.97	6.93	*
Scaled dula_AD[F]	0.59	0.38	.
Scaled dula_AD[M]	0.01	0.26	
Scaled dula_RD[F]	0.34	0.32	
Scaled dula_RD[M]	0.41	0.28	.
BS degree above[F]	0.08	0.87	
BS degree above[M]	-1.24	0.65	*

Hierarchical Poisson

	Mean	Standard Deviation	Sig.?
Years Driving [F]	-0.05	0.03	.
Years Driving [M]	-0.06	0.02	*
Annual Mileage [F]	0.87	0.55	*
Annual mileage[M]	1.77	0.56	*
Past violations[F]	-0.60	0.29	*
Past violations[M]	0.06	0.22	
sd.v	1.28	0.19	*
sigma2.v	1.71	0.59	*
tau.v	0.65	0.22	*

Comparing Parameters Across Collision Types

- One way to test whether a factor or event is a surrogate is to examine parameter magnitudes, significance and sign for a multitude of incident, near crash and crash types
  - For example, if near crashes and incidents are evaluated as surrogates, comparing driver, context and event related parameters in terms of their magnitude, significance and sign may be useful
  - Alternatively, if precipitating events or in-vehicle alerts are evaluated as additional surrogates, comparing their parameters may be useful

One Statistic For Comparing Parameters

- Likelihood ratio
- Compare the likelihoods of restricted and unrestricted specifications
- For a given X vector, the degrees of freedom equals K if K parameters are compared
- $-2(\text{Log-Likelihood}_{\text{restricted}} - \text{Log-Likelihood}_{\text{unrestricted}})$ for K degrees of freedom is chi-squared distributed
- A significant χ^2 indicates a statistical difference