Estimating Crash Risk Using Naturalistic Driving Study Data

Feng Guo, Ph.D.
Virginia Tech Transportation Institute
Department of Statistics
Virginia Tech

September 1st 2010

The Second International Naturalistic Driving Symposium

Estimating Crash Risk: Outline

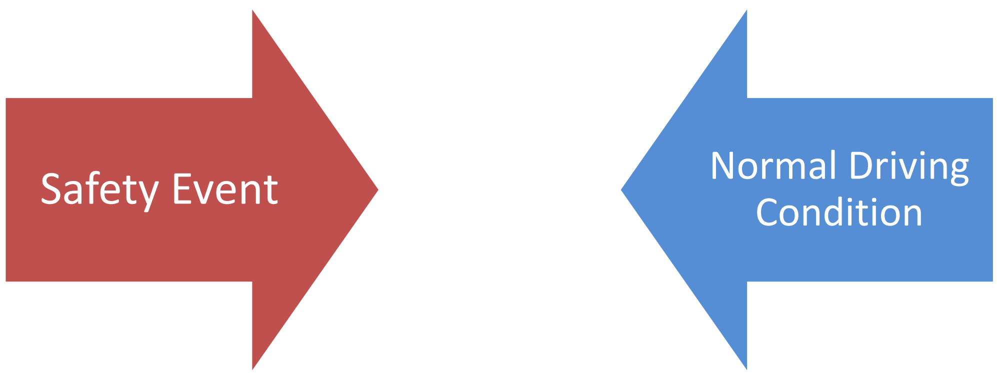
- Overview
- Case-cohort and case-crossover approaches: results and lessons learned
- Near-crashes as crash surrogates for risk assessment purpose

Naturalistic Driving versus Crash Database

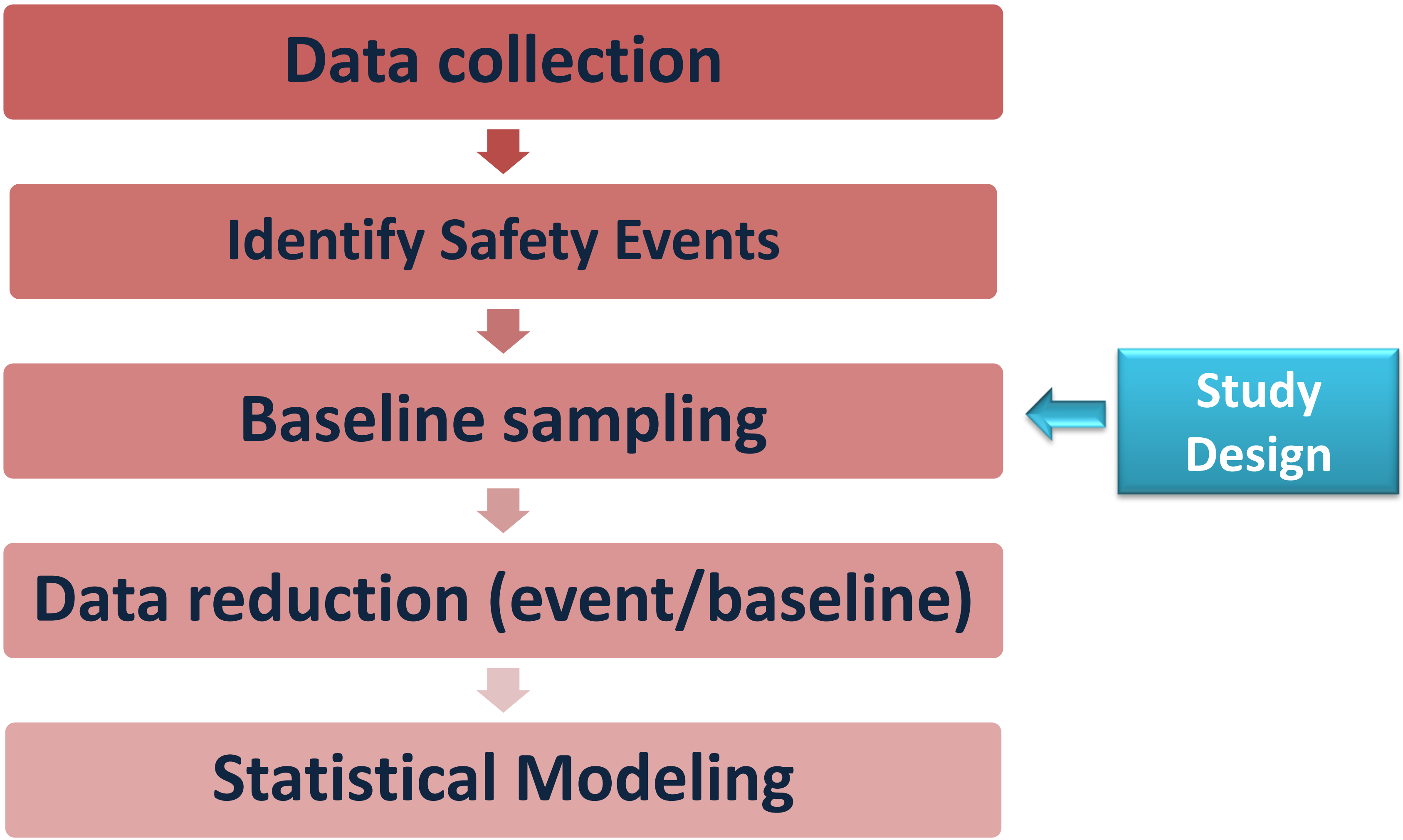
	Crash Database	Naturalistic driving data
Sample population	All drivers: limited selection bias	Participants only: selection bias
Source	Police report: small proportion of actual accidents	Data collection system: all safety events
Information source	Driver/witness statements, retrospective: information bias	High resolution video and instrument recording
Driver behavior	Limited/unreliable information	Accurate/detailed information through data reduction

Risk Assessment

- Presence of a factor at crash ≠ Risk
- Comparing exposure status for safety events and for normal driving conditions.
- Naturalistic driving data provides detailed and accurate exposure information

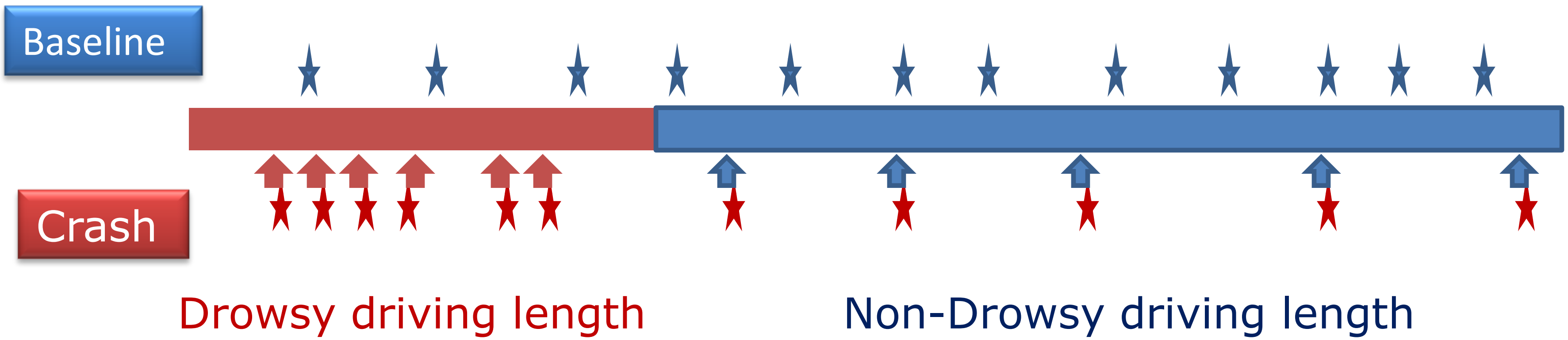


Modeling Crash Likelihood Framework

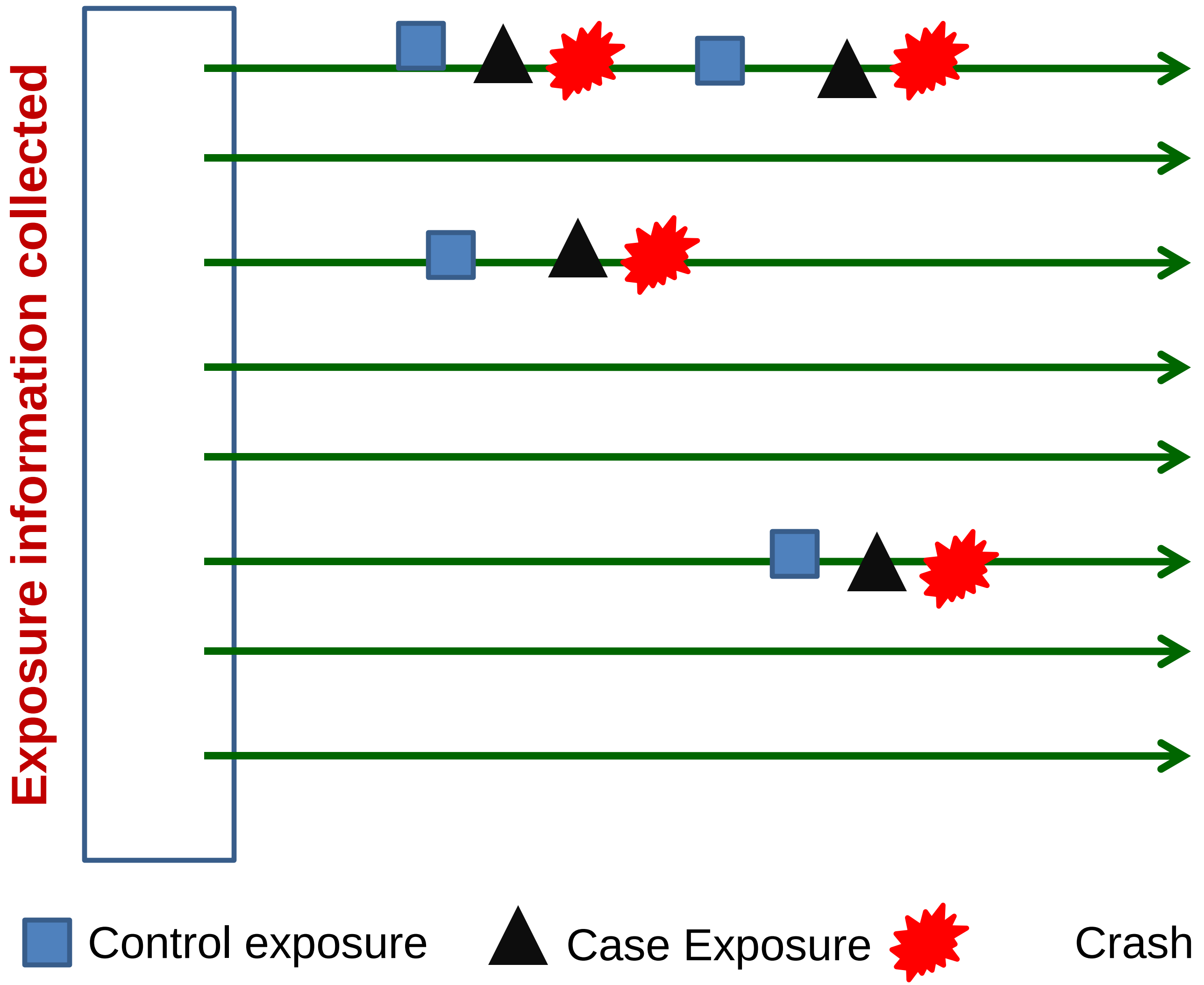


Modeling 100-Car: Case-Cohort Approach

- Sample short (6 second) epochs from the videos
- Sampling Scheme: Random sampling stratified by vehicle travel time
- Independent of crash/near-crash



Case-Crossover



Sample exposure immediate before crashes

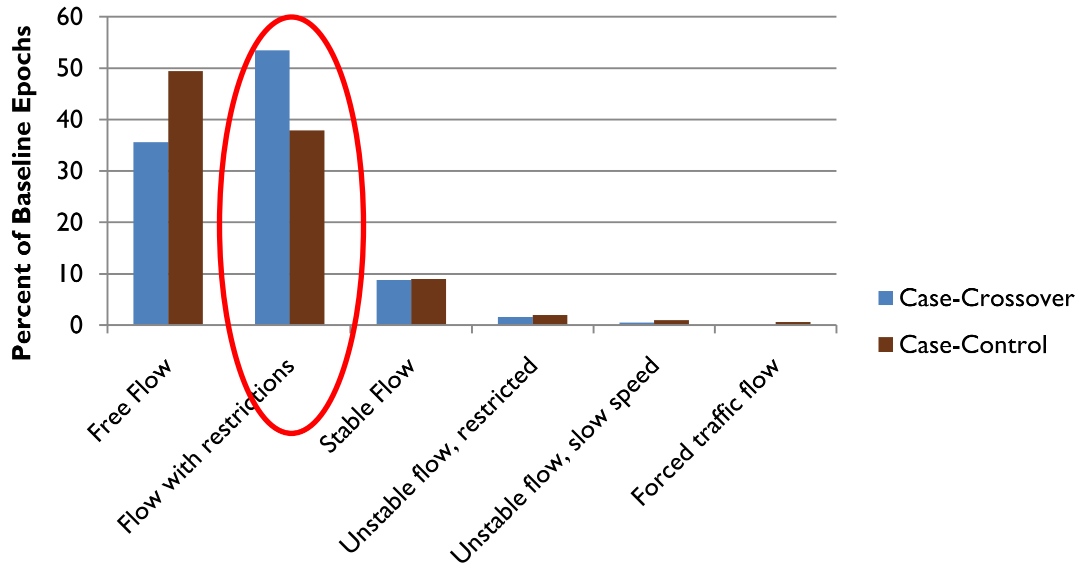
Sample exposure for time interval some period before crash

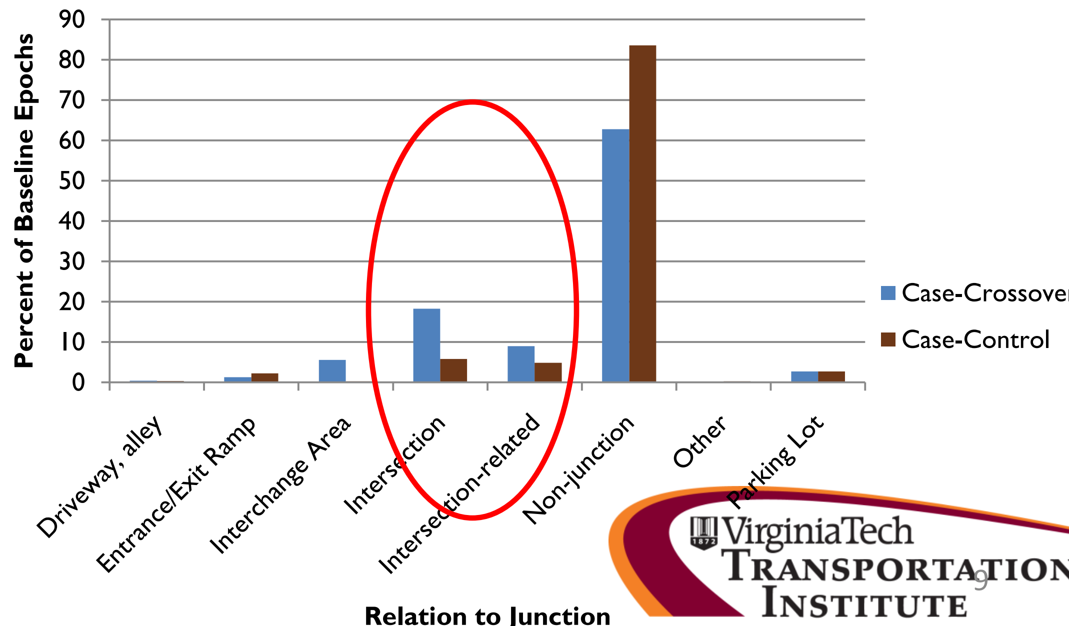
Matched Factors

- Driver ID
- Day of week (weekday versus weekend)
- Time of day (± 2 hours)
- Same GPS Location (± 100 Meters OR match to relation to junction)
- Must occur *prior to* crash/near-crash occurrence.

- Goal: 15 baselines for every crash/near-crash event.

Differences between the Case-Crossover and Case-Control

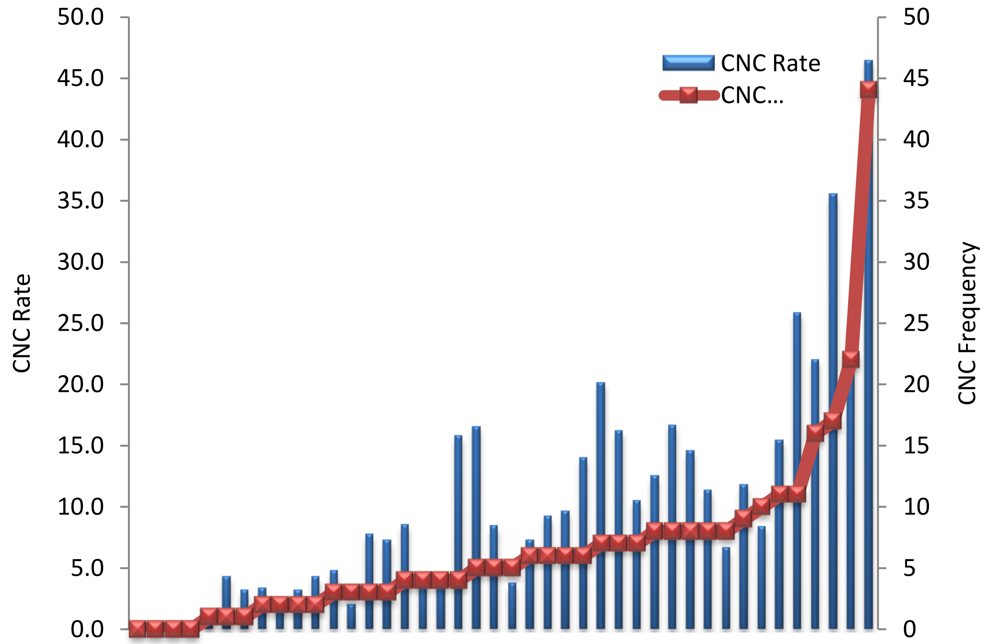




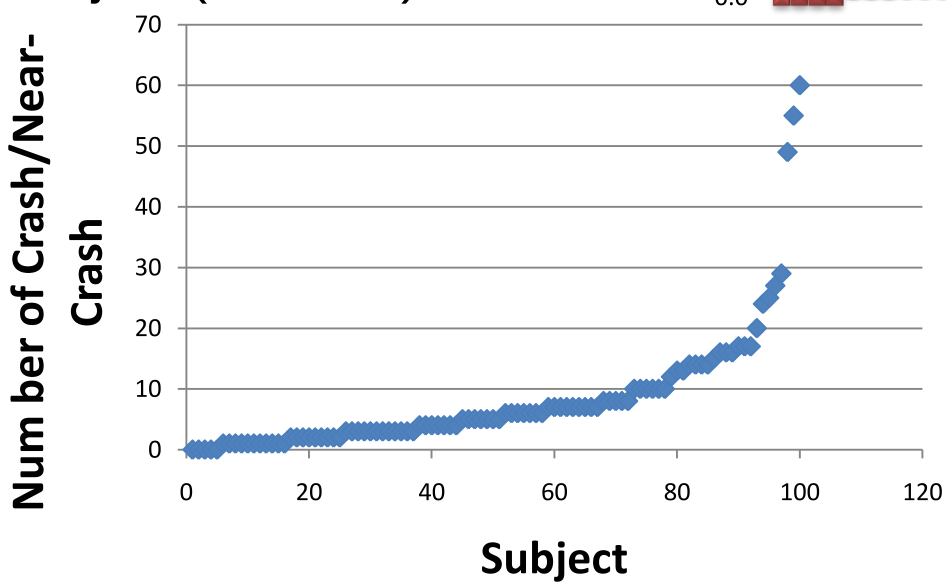
VirginiaTech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Individual Variation: Good Driver, Bad Driver

Teen driver crash/near crash rate



Number of safety events by subject (100-Car)



VirginiaTech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Case-Cohort: Generalized Mixed Effect Model

Model specification

$$y_i = \begin{cases} 1 & Crash \\ 0 & No\ Crash \end{cases}$$

$$y_i \sim Binomial(1, p_i)$$

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{\alpha}_i$$

where p_i is the probability of crash for *i*th observation
X_{1i} is the 1st covariate for event *i*;
β's are the regression parameters
$\boldsymbol{\alpha}_i$ is the driver specific random effect

Virginia Tech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Case-Crossover: Conditional Logistic Regression

Let p_{ij} be the probability of crash/near-crash for *jth* observation in *ith* matched set.

Define

$$Y_{ij} = \begin{cases} 1 & \text{if the jth observation in ith matched set is a crash /near} - \text{crash.} \\ 0 & \text{if the jth observation in ith matched set is a baseline.} \end{cases}$$

The matched sampling mechanism leads to:

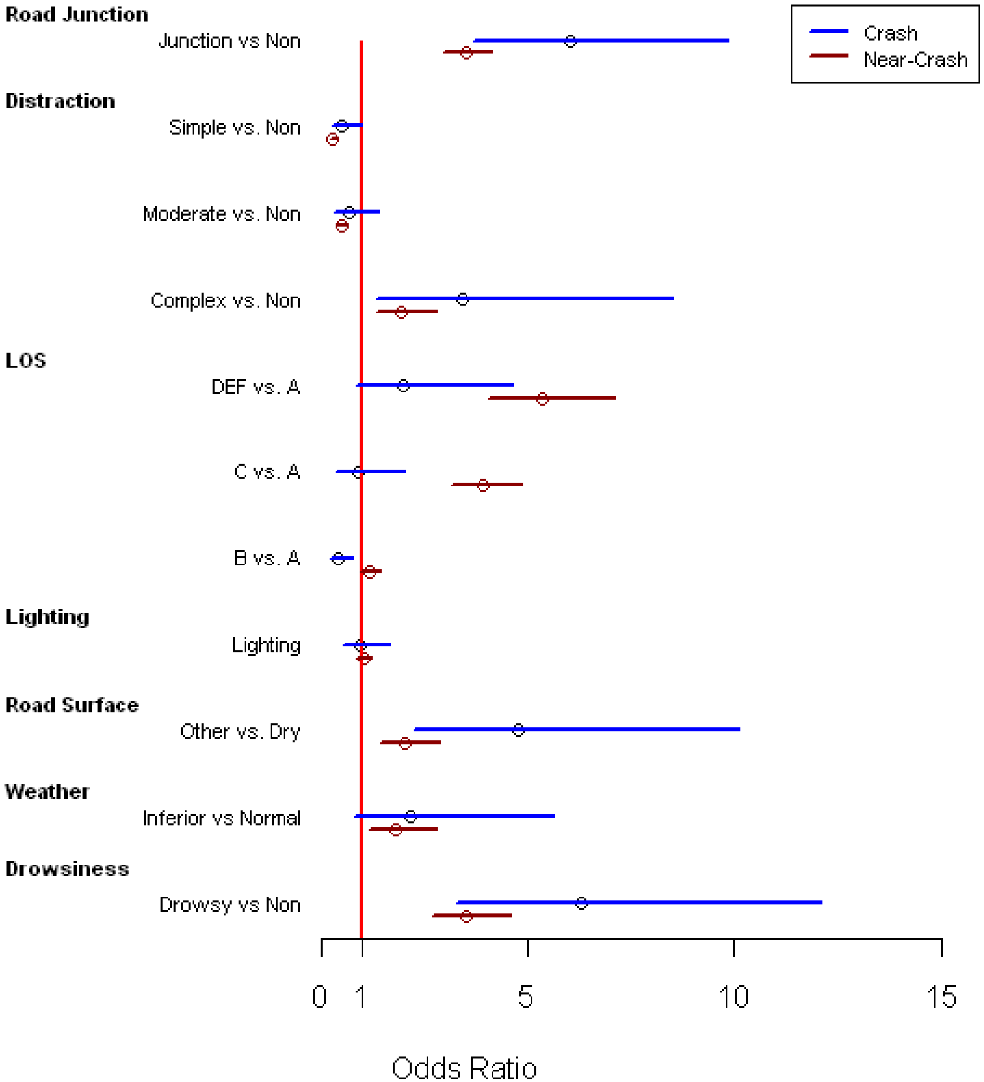
$$\sum_j Y_{ij} = 1$$

$$\text{logit}\left(p_{ij} \middle| \textstyle\sum_j Y_{ij} = 1\right) = \beta * \text{drowsy}_{ij},$$

In this model exp (β) is the estimated OR.

Case-Cohort: Crash Risk

- Drowsiness increases crash risk by 6 times
- Complex secondary task increase the risk by 3 times
- Crashes are more likely to happened in roadway junction areas (6-fold increase)
- Crashes are 5 times more likely to happened on wet, snowy, or muddy road surface.



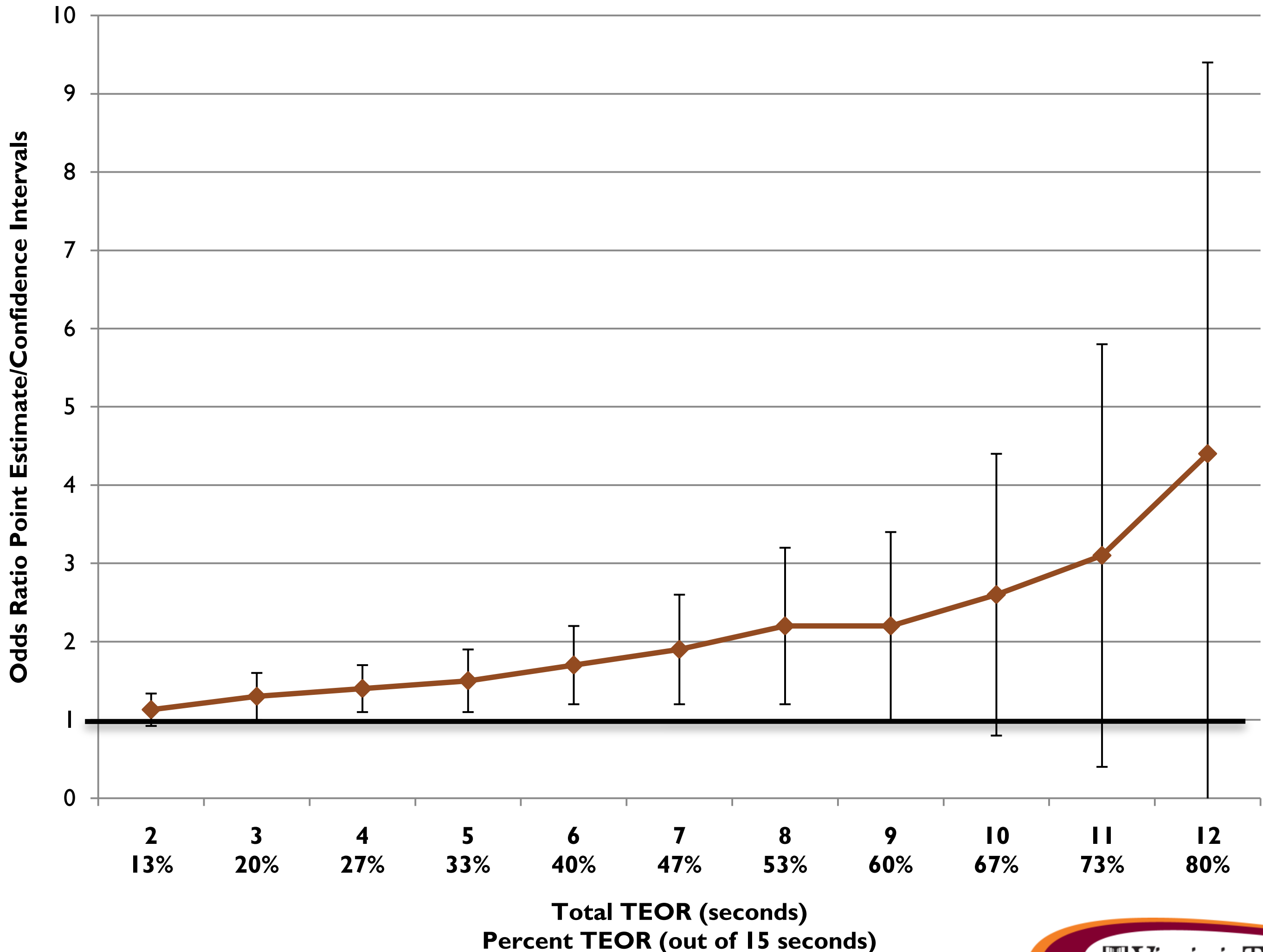
INSTITUTE
DRIVING TRANSPORTATION WITH TECHNOLOGY

Comparing Secondary Task Engagement OR for Case-Crossover versus Case-Control

Distraction	Case Cross-over Odds Ratio	95% Odds Ratio Confidence Limits		Case-Control Odds Ratios	95% Odds Ratio Confidence Limits	
Simple	0.8	0.62	1.05	1.2	0.88	1.57
Moderate	**1.3**	**1.00**	**1.70**	**2.1**	**1.62**	**2.72**
Complex	**2.1**	**1.19**	**3.58**	**3.1**	**1.72**	**5.47**

VirginiaTech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Crash Risk Increase Monotonically with Total Eyes Off Forward Roadway



VirginiaTech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Comparison of Case-Cohort and Case-Crossover

- Case-Cohort: Greater generalizability to not only driver behavior but also environmental and roadway risk assessment.
- Case-Cohort: Simpler to conduct and less resource intensive.
- Case-Crossover: Greater precision as potential confounding factors are controlled through baseline sampling.

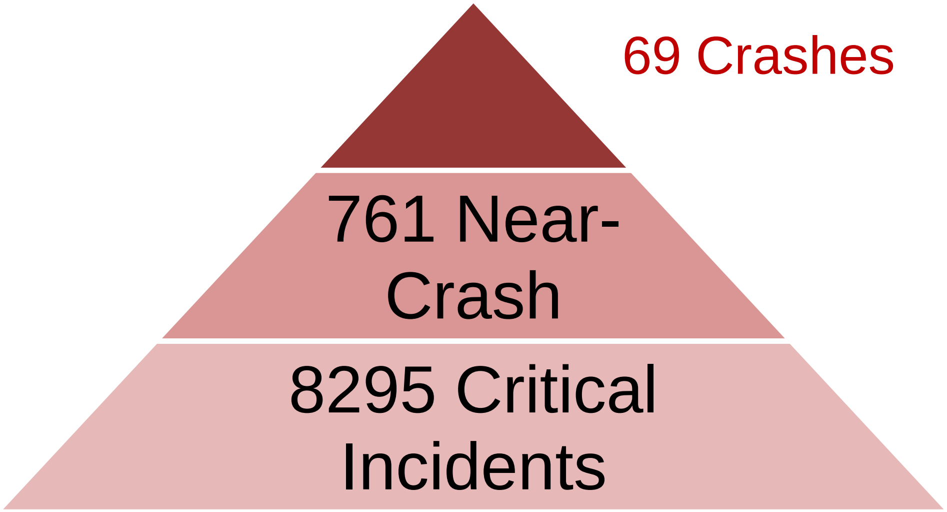
VirginiaTech Transportation Institute
DRIVING TRANSPORTATION WITH TECHNOLOGY

Can Near-Crashes Serve as Crash Surrogates?

- They are different by definition!

- It depends on the purpose the study.
  - This analysis focuses on the impacts for risk assessment purpose.

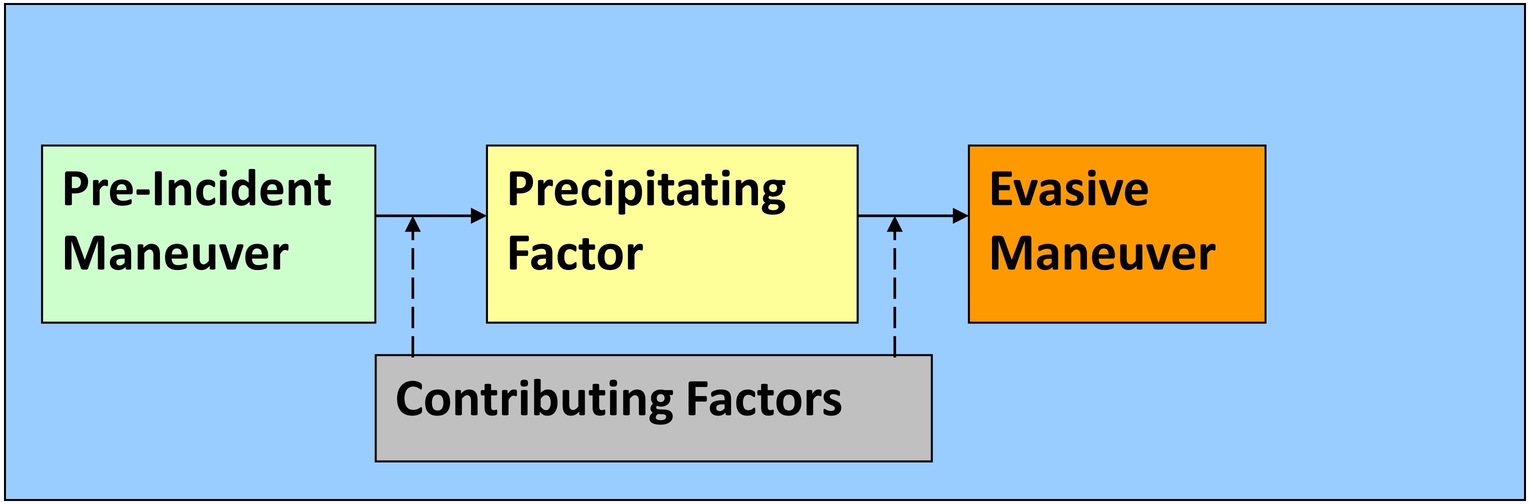
Crash Surrogate



- Less-severe events happen more frequently than severe events
- Severe events can be reduced by reducing less sever events

1. The causal mechanism for surrogates (near-crashes) and crashes are the same or similar.
2. There is a strong association between the frequency of surrogate measures and crashes under different settings.

Driver Reaction on Crash and Near-crash



All Conflict Types | **Conflict with Leading Vehicle**

i	Crash	Near-Crash		Crash	Near-Crash
Reaction	45	723	Reaction	5	377
No-Reaction	23	37	No reaction	9	0
Perc. Reaction	**66%**	**95%**	Perc. Reaction	**36%**	**100%**

Frequency Relationship

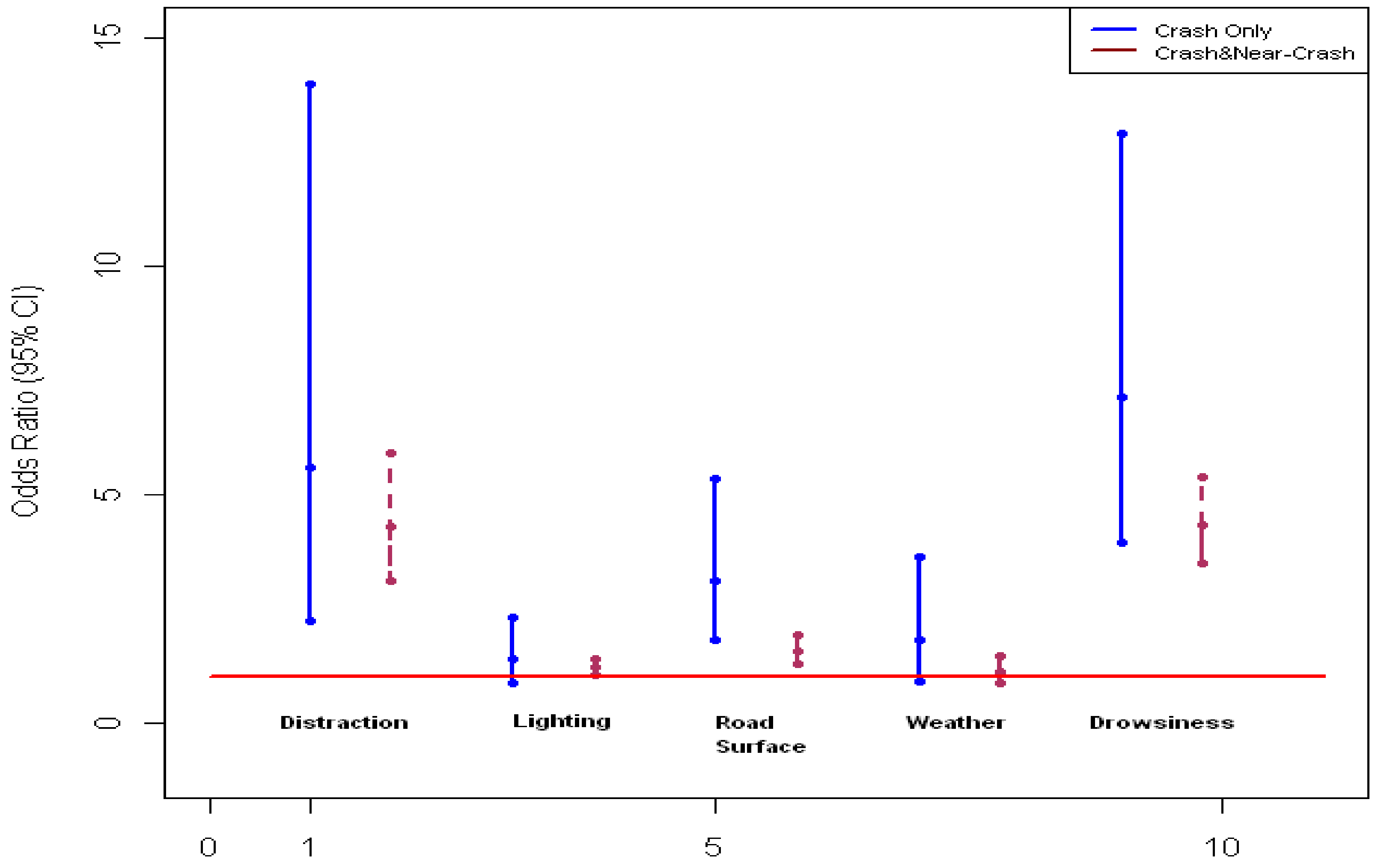
	Constant Crash to Near-Crash Ratio		Measure for Association	
Factors	p-value	Significant	R-squared	Adjusted R^2
Gender	0.26	NO	NA	NA
Age Group	0.23	NO	**0.91**	**0.87**
Level of Service (LOS)	<0.001	YES	**0.5 (0.72*)**	**0.33 (0.45*)**
Lighting Conditions	0.414	NO	**0.97**	**0.95**
Road Alignment	0.02	YES	**0.99**	**0.99**
Road Surface Condition	0.02	YES	**0.99**	**0.99**
Weather	0.32	NO	**0.99**	**0.99**

$$y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i$$

	Coefficient	***p*-value**
Intercept	-2.31	<.0001
Near-Crash	**0.21**	**<.0001**

Sensitivity Analysis



Surrogate Measure: Summary

- Using crashes plus near-crashes will lead to a conservative but more precise result in risk assessment.
- For smaller studies with an insufficient number of observed crashes, there is a definite benefit to using near-crashes as a crash surrogate.
- Caution should be used when interpreting the results of risk evaluation.