UMTRI

A Multivariate Analysis of Crash and Naturalistic Event Data in Relation to Highway Factors Using the GIS Framework

Update

Third SHRP2 Safety Research Forum

July 17, 2008

T. Gordon, P. Green, L. Kostyniuk, M. Barnes, S. Bogard, D. Blower, D. LeBlanc

UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE

Objective

Capture common elements in how highway factors are associated with

- **Crashes (as recorded in crash data)**
- **Driving behaviors (as recorded by naturalistic driving data from FOT)**
- **Focus**
  - **On spatially referenced databases and GIS tools**
  - **On mechanisms, not just statistical association**

UMTRI

Definitions

- **Control - effectiveness of tactical and operational aspects of the driving task**

  (acquire and track reference information for speed and steering adjustment)

- **Disturbed control - any interruption or delay in the process of**
  - **perception (seeing lane edge or other boundary features)**
  - **recognition (filtering of the road scene)**
  - **judgment/decision (of steering, throttle or brake pedals)**
  - **action (apply corrections)**

UMTRI

Assumptions

- Single vehicle road departure crashes occur primarily under conditions of disturbed control
- Naturalistic driving data contain measurable episodes of disturbed control
- Crash surrogates exist and are based on a combination of
  - objective measures of disturbed control (from on-board sensors)
  - highway geometry factors
  - off-highway factors
  - environmental factors
- Crash surrogates can be related to actual crashes

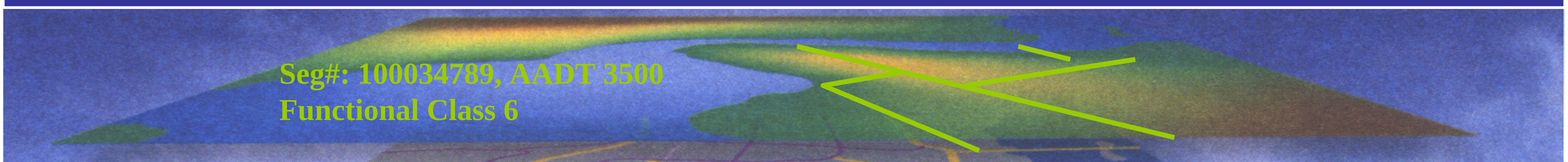
Approach

- Identify candidate surrogates - related to disturbance of control
- Test if candidate good measure of crash risk – (Bayesian multivariate generalized models)
- Test if candidate indicator of extremes in process that includes crashes in limit (extreme value theory)
- Validate and replicate

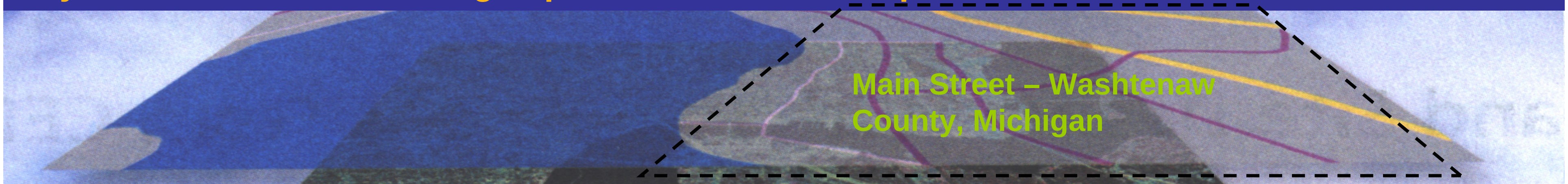
Spatially Referenced Data

Layer One - SE Michigan Road Departure Crashes (2001-2005) (Lat/Long)

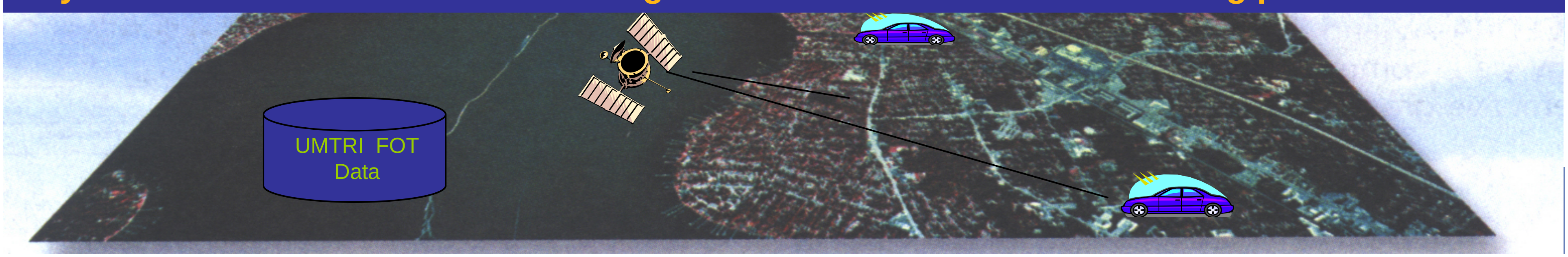
Layer Two - Michigan 2005 HPMS- Table Data



Layer Three - State of Michigan public roads base map



Layer Four - UMTRI Naturalistic driving data from FOT fleets with Lat Long positions



Crash Data

- Road Departure crashes from MI Vehicle Crash Data file 2001–2005

- 71,308 crashes in SE MI 8-county area
  - Direction of travel identified
  - Direction of road departure
    - 15,187 left departures
    - 23,836 right departures
    - 32,287 unknown
  - 50,654 crashes located to HPMS segments
- Driver - age, sex
- Environmental conditions (day/night, wet/dry pavement)

Exposure - Crash rate

(crashes in directional segment) divided by (directional 5-yr volume* segment length)

Road Segment Data

Our Universe - HPMS road segments with direction - traversed by FOT vehicles in SE MI 8-county area – (5,965 directional segments)

- **From HPMS**
  - **Segment length**
  - **AADT**
  - **Functional class**
  - **Area type – Urban, Rural**
  - **Number of thru lanes**
  - **Shoulder type***
  - **Shoulder width right***
  - **Shoulder width left***
  - **Speed limit***

***Sample data – involves imputation**

- **From FOT data**
  - **Curve presence**
  - **Radius of curvature**
  - **Average change of heading**
  - **Available lateral maneuvering distance**

Candidate surrogates

from 220,000 vehicle miles FOT data

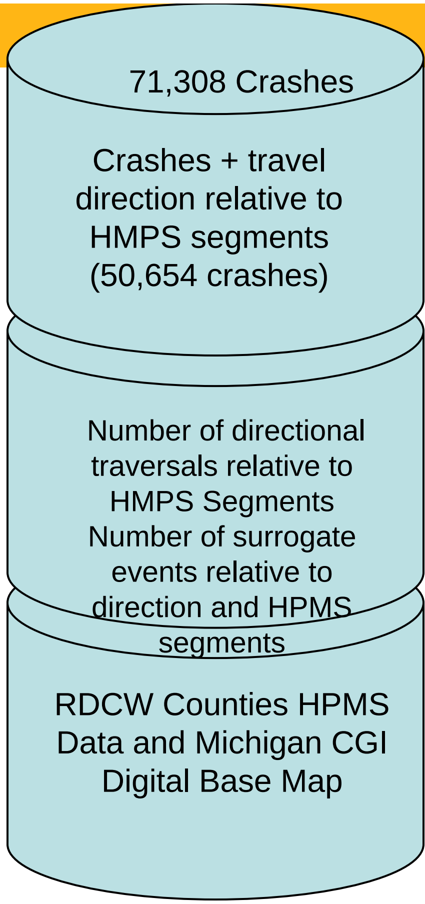
- **Imminent lane departure warning (LDW Alert)**
- **Approach speed too fast for curve (curve speed warning)**
- **Inverse time to lane crossing**
- **Peak lateral deviation in lane**
- **Peak projected lateral deviation**
- **Driver looking away for 2 sec+**
- **Coherency between tracking error and steer response**
- **Steering rate less than small threshold for at least 4 seconds**
- **Others from search for benchmarks of disturbed control**
  - **Cell phone use**
  - **Yaw deviation associated with discontinuity in lane markings**

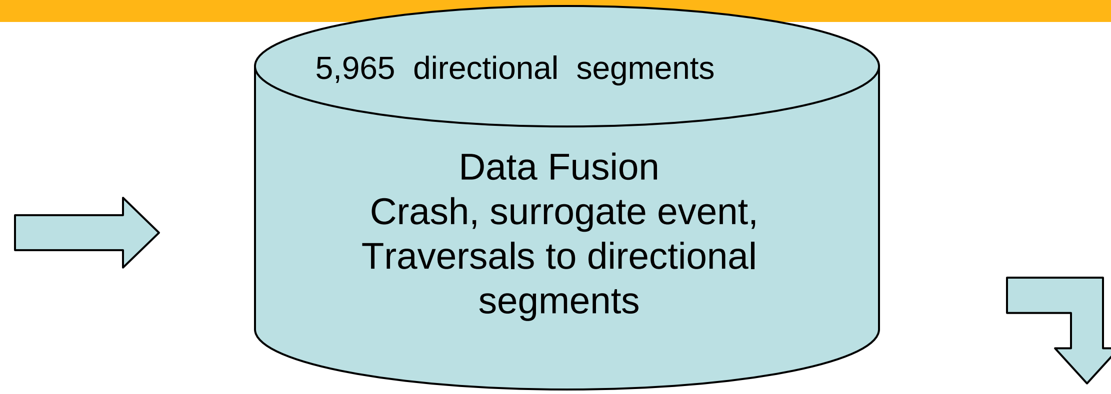
Exposure - Surrogate event rate
(Events in directional segment) divided by (number of FOT traversals * segment length)

Other information in FOT data

- **Driver age, sex**
- **Environmental conditions – light, windshield wipers on/off**

UMTRI





Data Fusion Crash, Alerts, Crash Rate, Alert Rate, and Traversals by HPMS Segment					
HPMS Segment	Crashes		LDW Alert Events		
HPMS Seg and Direction	Number of Crashes (Located within 75 Feet of HPMS Segment)	Crash Rate* per 1M Entering Vehicles	Number of Alerts (Located within 75 Feet of HPMS Segment)	Number of Traversals	Alert Event Rate*
500003970216-1	19	0.495094	2	31	0.029100645
630020410414-1	7	0.19936	NULL	97	#VALUE!
580000190111-2	8	0.083505	NULL	16	#VALUE!
810000190111-1	7	0.102554	1	44	0.024975025
500006440417-1	4	0.255788	NULL	34	#VALUE!

Data Fusion Crash, Alerts, Crash Rate, Alert Rate, and Traversals by HPMS Segment													
HPMS Descriptors													
HPMS Sampled Data								HPMS Universe Data					
CurveA (0,1)	CurveB (0,1)	CurveC (0,1)	CurveD (0,1)	CurveE (0,1)	CurveF (0,1)	Shoulder Type	Shoulder WidthR	Shoulder WidthL	AADT	Functional Class	Segment Length	Thru Lanes	Urban Rural
0	0	0	0	0	0	3	4	0	9485	16	2.217	2	4
1	0	0	0	0	0	6	0	0	71523	14	0.269	4	4
1	0	0	0	0	0	4	12	10	68800	11	0.763	6	3
1	0	0	0	0	0	4	10	10	41100	11	0.91	4	4
0	0	0	0	0	0	6	0	0	6855	17	1.25	2	4

UMTRI

Candidate Surrogate LDW Alert

- Predicted vehicle path will cross solid lane boundary (edge or center line) or dashed line boundary into occupied space
  - Conditions:
    - Vehicle speed >25 mph
    - Restriction on high steering rate in past 5 seconds
    - No turn signal in the past 5 sec
    - No brake in the past 5 sec
    - Not on local street
    - Actual tracking on boundary to be crossed

Bayesian Multivariate Generalized Model

Crash Data

Response	Crash Events
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

FOT Data

Response	Surrogate Events
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

Bayesian GMV model
Surrogate Candidate - LDW Alert

Response Variables

Number of crashes
Number of LDW alerts

Exposure Variables

AADT, traversals, segment length

Explanatory Variables

Curve – Yes, No
Freeway – Yes, No
Area – Rural, Urban
Right Shoulder – Yes, No

Exploratory Analysis

Fit Two Separate Poisson Models for Rates

Parameter	Estimates Crash Model	 Alert Model
Intercept	8.31	9.78
Curve	-0.39	-0.56
Freeway	-5.90	-6.69
Area	-3.81	-3.08
Shoulder	-7.66	-7.35
Curve*Freeway	-0.22	-0.17
Curve*Area	1.27	1.83
Curve*Shoulder	-0.17	0.16
Freeway*Area	-0.29	-1.84
Freeway*Shoulder	6.30	6.45
Area*Shoulder	1.55	1.05

Seemingly Unrelated Regressions (SUR)

Zellner (JASA, 1962) describes a method for estimating the parameters in a set of regression equations called Seemingly Unrelated Regressions (SUR).

$$Y_1 = X_1\beta_1 + \varepsilon_1$$
$$Y_2 = X_2\beta_2 + \varepsilon_2$$

Seemingly Unrelated Regressions (SUR)

As a system of equations, SUR may be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

crashes

alerts

Or as a new linear model

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \Sigma)$$

← Random component

$$Var(\varepsilon) = \Sigma = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I \\ \sigma_{21} I & \sigma_{22} I \end{bmatrix}$$

← Correlation structure

UMTRI

Seemingly Unrelated Regressions (SUR)

Properties of SUR:

1) **The system satisfies the properties of Weighted Least Squares (WLS)**

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

2) **A correlation structure is induced between Y_1 and Y_2**

3) **Tests of hypotheses may be conducted**

$$H_0 : \beta_1 = \beta_2$$

Bayesian SUR for Log-Linear Models

Likelihood:
(After suitable transformations)

$$Y_1 \sim \text{Normal}(X_1\beta_1, \Sigma_1)$$ **crashes**

$$Y_2 \sim \text{Normal}(X_2\beta_2, \Sigma_2)$$ **surrogates**

Or as a system:

$$Y \sim \text{Normal}(X\beta, \Sigma)$$

Vague Prior:

$$\beta \sim N_p(0, 10^6 I)$$

Σ Fixed covariance matrix estimated from ordinary least squares (OLS) residuals

Relative Risks

- **Use Relative Risk (RR) as exposure-based risk measure**
- **RR is a ratio of rates**
- **Test if there is a difference in RR between crashes and surrogates**
- **If log RR difference is zero, conclude that surrogate is a good surrogate for crashes**

UMTRI

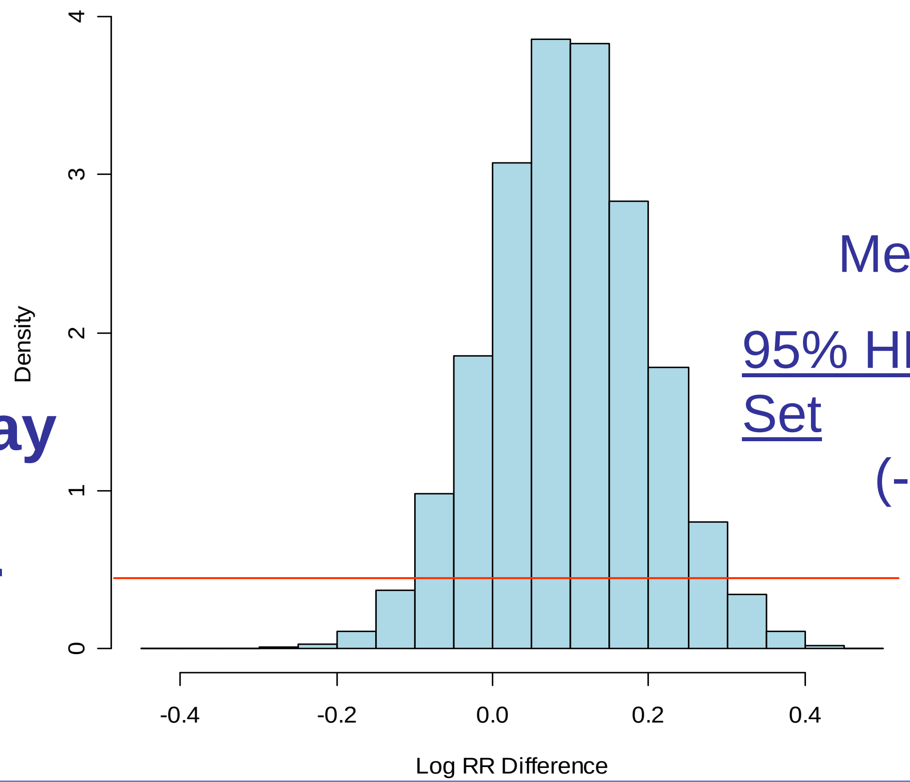
LDW Alerts as Good Surrogate

Estimated Difference in Log Relative Risks (RR)
(Histogram of 30,000 Random Variables Simulated from the Posterior Distribution of the Log RR Difference)

Comparing:

Curve to
No Curve

Not On Freeway
In Rural Area
With Shoulder



Mean: 0.1

95% HPD Credible Set
(-0.1, 0.3)

UMTRI

Next Steps

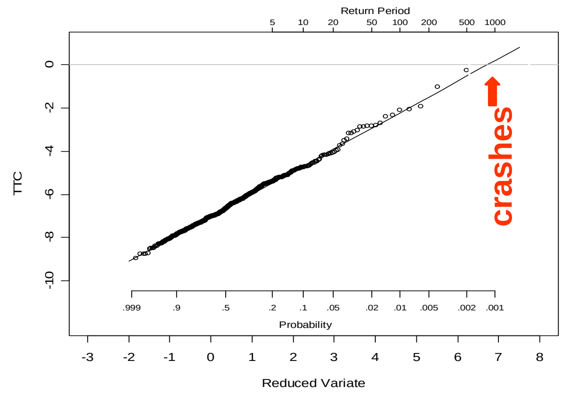
- Other candidate surrogates
- Other geometric variables
- Driver variables – age, sex
- Environmental variables – day/night, rain
- Use promising candidates in extreme value analysis to predict crash trend and estimate relative risk.
- Replicate (to extent feasible) with other data – VTTI.

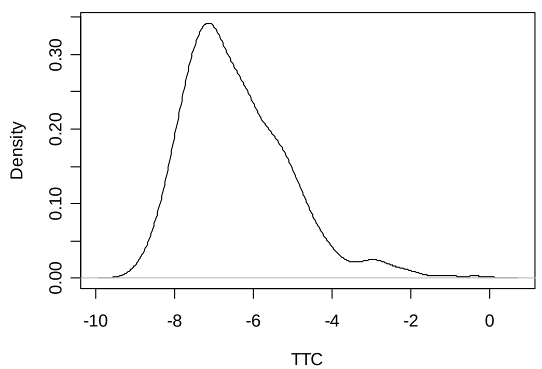
Go Blue

Thank You !

Promising candidate surrogate
Extreme value analysis
Predict crash trend and estimate relative risk

- **Crash is a rare event, occurs at very extreme value of a surrogate**
- **Use EVT to model rare events that lie outside the range of available observations**
- **Select smallest observation from each of many samples – obtain extreme value distribution**
- **Can estimate number of extreme observations before seeing one crash**
- **Model rare event - by GLM, random effects model , fully Bayesian hierarchical models in regression setting**
- **Can control for driver characteristics, road type, weather, etc.**





UMTRI