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

University of Michigan
Transportation Research Institute
Aim of project

Capture common elements in how highway factors are associated with

- Crashes (as recorded in crash data)
- Driving behaviors (as recorded by FOT data)
Control is defined here as the effectiveness of tactical and operational aspects of the driving task.

Crashes occur when there is a loss of control (for whatever reason).

Not every loss of control event results in a crash.

Road departure crashes.

FOT data – contain indicators of control problems:
- Certain lane departure warning
- Others - to be determined
Hypotheses

- Crash surrogates exist and can be found in FOT data
- The crash surrogates are related to highway features and to driver-related factors
- Crash surrogates are related to control
Data

- 71,000 relevant crash events in SE Michigan from 2001-2005 crash data
- 220,000 miles FOT data in SE Michigan
- 8,300 road departure warnings (~0.1 per vmt)

Roadway information
  - Michigan Base GIS
  - HPMS
Currently 4 sources of data are spatially referenced either using existing GPS data or completing spatial joins using GIS software utilities.

Layer One - Michigan Run-Off-Road Crashes-Spatial Data (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 including Lat Long Positions
These sources are “joined” to create a flat file for analysis of research questions.

<table>
<thead>
<tr>
<th>Functional Class</th>
<th>FULL ID</th>
<th>BPT</th>
<th>EPT</th>
<th>GPSLat</th>
<th>GPSLong</th>
<th>speed m/s</th>
<th>csw alert level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>181518</td>
<td>61012483</td>
<td>61012034</td>
<td>42.20582</td>
<td>-83.6874</td>
<td>17.77773</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>181610</td>
<td>61012483</td>
<td>61012034</td>
<td>42.20603</td>
<td>-83.6874</td>
<td>18.33332</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>259331</td>
<td>61012305</td>
<td>61011806</td>
<td>42.20664</td>
<td>-83.6868</td>
<td>21.94439</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>259332</td>
<td>61012305</td>
<td>61011806</td>
<td>42.20734</td>
<td>-83.6883</td>
<td>25.55554</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>259578</td>
<td>61012305</td>
<td>61011806</td>
<td>42.20765</td>
<td>-83.6868</td>
<td>25.77771</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>259597</td>
<td>61012372</td>
<td>61012437</td>
<td>42.20637</td>
<td>-83.6855</td>
<td>31.38888</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>259598</td>
<td>61012372</td>
<td>61012068</td>
<td>42.21036</td>
<td>-83.6853</td>
<td>34.72214</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>259722</td>
<td>61012458</td>
<td>61012372</td>
<td>42.21675</td>
<td>-83.5461</td>
<td>12.22219</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>259804</td>
<td>61012468</td>
<td>61012191</td>
<td>42.21771</td>
<td>-83.5474</td>
<td>19.72218</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>260510</td>
<td>61012669</td>
<td>61012672</td>
<td>42.21778</td>
<td>-83.5477</td>
<td>20.27773</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>260591</td>
<td>61012676</td>
<td>61012709</td>
<td>42.21784</td>
<td>-83.5468</td>
<td>23.61110</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>260730</td>
<td>61012669</td>
<td>61012677</td>
<td>42.21603</td>
<td>-83.5579</td>
<td>33.61110</td>
<td>1</td>
</tr>
</tbody>
</table>
Additional sources of spatial data being collected from selected agencies include:

- intersections
- signalized intersections
- traffic control signage locations
- aerial photographs-pavement condition data

IRI, PR Rating “X”.
Seg No. 0025946590
Overlap of FOT driving locations and road departure crashes

- 37,117 locations+ direction where FOT driving data overlap road departure crash
- 22 locations with 10+ road departure crashes within 20 meters of each other
- 1,654 locations with 3+ road departure crashes within 20 meters of each other
Road departure crashes

- 3+
- 10+
Road departure crashes

- 3+
- 10+
Sites with FOT data and at least 3 road departure crashes
Experimental Design

Highway Related Factors

Environment
Weather, lighting, traffic volume

CRASH
NON CRASH

FOT – control
“bad” ---------- “good”
Crash/Non Crash events

- Non crash
  - Road Segment with no crashes
    - VMT, geometric features

Examine crash/non crash events
  - Literature review
  - Clustering of crashes at locations
  - Patterns in highway factors
  - What is relevant in road departure crash
Control metrics

- Identify discriminators of “good” “bad” control
- What are the patterns of control indicators at crash and non crash locations?
  - Discrete, Continuous, dynamic sequence?
- Can the patterns of control indicators be used to predict high crash locations?
- Do they satisfy surrogate evaluation criteria?
Driver Factor Questions

- Can relevant driver factors be identified?
  - Aggressive driving
  - Distracted driving
  - Fatigued driving
  - Engaged driving

- Can their influence be controlled for in analysis?
Challenges for Year 1

- Refine definitions and hypotheses
- Experimental design
- Tests of reasonableness
- Supporting statistical models and analyses methods
Crash Data and Models

Standard models are available for assessing risk in crash data

Example: Hierarchical Bayesian Models

<table>
<thead>
<tr>
<th>Explanatory Factors</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver</td>
<td>Crashes</td>
</tr>
<tr>
<td>Environment</td>
<td>(Exposure)</td>
</tr>
<tr>
<td>Vehicle</td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td></td>
</tr>
</tbody>
</table>

Likelihood: Poisson
First Prior: Gamma or Normal (attach mean to log-linear model with explanatory variables)
Second Prior: noninformative proper priors on model parameters
Crashes and Surrogate Measures

But this project involves an additional component. Can surrogates be related to actual crashes?

<table>
<thead>
<tr>
<th>Crash Data</th>
<th>FOT Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response</strong></td>
<td><strong>Surrogates</strong></td>
</tr>
<tr>
<td>(Exposure)</td>
<td>(Exposure)</td>
</tr>
<tr>
<td><strong>Explanatory Factors</strong></td>
<td><strong>Explanatory Factors</strong></td>
</tr>
<tr>
<td>Driver</td>
<td>Driver</td>
</tr>
<tr>
<td>Environment</td>
<td>Environment</td>
</tr>
<tr>
<td>Vehicle</td>
<td>Vehicle</td>
</tr>
<tr>
<td>Highway</td>
<td>Highway</td>
</tr>
</tbody>
</table>
A Multivariate Generalized Model

\[ r_{i1}, r_{i2} \mid \beta_1, \beta_2, \Sigma_i \sim N_2 \left( \begin{bmatrix} x_{i1}^T \beta_1 \\ x_{i2}^T \beta_2 \end{bmatrix}, \Sigma_i \right) \quad i = 1, \ldots, N \]

**log crash rate:** \[ r_{i1} = x_{i1}^T \beta_1 + \epsilon_{i1} = \beta_{10} + \beta_{11} x_{i11} + \ldots + \beta_{1p} x_{ilp} + \epsilon_{i1} \]

**log surrogate rate:** \[ r_{i2} = x_{i2}^T \beta_2 + \epsilon_{i2} = \beta_{20} + \beta_{21} x_{i21} + \ldots + \beta_{2q} x_{i2q} + \epsilon_{i2} \]

Relative risks: \[ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \]

Crashes \quad Surrogates

Cov (\epsilon_{i1}, \epsilon_{i2}) = \Sigma_i

A variance stabilizing transformation is available

\[ H_0 : \beta_1 = \beta_2 \quad \text{excluding intercepts} \quad \beta_{10}, \beta_{20} \]

\[ H_0 : \beta_{11} = \beta_{21} \]
Advantages of the Multivariate Generalized Model

- WLS can be used to estimate ML parameters in Poisson log-linear models. The proposed model is a multivariate generalization of WLS. (The marginal models for crashes and surrogates are equivalent to the usual univariate WLS models).
- The model incorporates a correlation structure between crashes and crash surrogates (through $\sum$).
- Overdispersion relative to the Poisson model is handled through the scale parameters in the Normal distribution.
- All calculations are tractable and fast. Models used in econometrics (no need for MCMC simulation).
- Formal hypothesis tests can be carried out to determine if surrogates are good measures of crash risk.
Extreme Value Theory

- Idea is to model rare events that lie outside the range of available observations.
- The smallest observation is selected from each of many samples.
- The resulting sample of minimum values is the sample of extreme values.

Example: In each time frame of 15 min, record the minimum time to road departure.

- EVT can also be approached in a Bayesian framework in a regression setting.
Extreme Value Theory

Return Period

Time to Lane Departure

Probability

Reduced Variate

crashes
Extreme Value Theory

![Graph showing the distribution of time to lane departure with density on the y-axis and time to lane departure on the x-axis. The graph has a peak at a time to lane departure of approximately -9.]

Time to Lane Departure

Density

0.00 0.10 0.20 0.30

-10 -8 -6 -4 -2 0
Extreme Value Theory

Return Period

Probability

Time to Lane Departure

Reduced Variate

Probability

5 10 20 50 100 200 500 1000

5 10 20 50 100 200 500 1000

-3 -2 -1 0 1 2 3 4 5 6 7 8

-10 -8 -6 -4 -2 0 2 4 6 8

.999 .9 .5 .2 .1 .05 .02 .01 .005 .002 .001

Reduced Variate
Thank You!