Investigating Lane-specific Traffic Parameters to Lane-changing Related Crashes

Zhi Chen, Graduate Research Assistant
Xiao Qin, Associate Professor
Mohammad Razaur Shaon, Graduate Research Assistant
Department of Civil and Environmental Engineering
South Dakota State University

August 22, 2014
Outline

• Introduction
• Objective
• Literature Review
• Methodology
• Data Collection and Processing
• Analysis
• Conclusions
Introduction

• Crashes related to lane changes are referred to as lane-changing related crashes, often presented as sideswipe crashes.
• Freeway management system has greatly increased the ability to collect, store, and process traffic data in a large-scale and real-time fashion through detection.
• Predicting near-term likelihood of crashes helps to identify hazardous conditions and develop preventive countermeasures to proactively address safety problems.
Objective

Using lane-specific and real-time traffic data to investigate the relationship between lane-changing related crashes and traffic flow parameters.
Literature Review

Lane change theory:

- Classified lane changes into free, forced, and cooperative actions based on the interaction between drivers (Hidas, 2005).
- Developed a hierarchical structure of the decision process to model lane changes on urban highways (Gipps, 1986).
- Investigated driver lane-changing behavior by using an instrumented vehicle. (Brackstone et al., 1998).
- Studied the effect of relative velocity between the target lane and subject lane and the spacing distribution on (discretionary) lane change decisions using Next Generation Simulation (NGSIM) data (Lee et al., 2013).
- Studied lane change duration based on the level of congestion on a freeway (Hill and Elefteriadou, 2013).
Real-time lane-changing related crash study:

- Identified Average Flow Ratio (AFR), the ratio of flows from one lane to its adjacent lane(s), as a statistically significant factor of lane change intensity through a video data study (Chang and Kao, 1991).
- Compared between the contributing factors for rear-end and lane-changing crashes using loop detector data (Lee et al., 2006).
- Estimated the likelihood of a lane-changing crash in the center lane compared to its left or right lane, respectively (Lee et al., 2009).
- Predicted lane-changing related crashes using neural network (Pande and Abdel-Aty, 2006).
Methodology

- **Matched case-control logistic regression** was used to investigate the effects of traffic flow variables while controlling for other non-traffic-flow parameters such as weather and geometric design elements.

- Each crash is considered as an *event*; non-crash events with similar non-traffic-flow variables are considered as *controls*, and each case and its controls constitute a *stratum*.

- Likelihood function:

\[
l(\beta) = \prod_{i=1}^{N} \exp \left( \sum_{k=1}^{K} \beta_{k} x_{ij} \right) / \sum_{j=0}^{J} \exp \left( \sum_{k=1}^{K} \beta_{k} x_{ij} \right)
\]

where \(x_{ijk}\) is the \(k\)th traffic flow variable for the event \((j=0)\) or the \(j\)th control in the \(i\)th stratum; \(\beta_{k}\) is the coefficient; \(i = 1, 2, ..., N; j = 0, 1, ...,\)
Data Collection

• A 62-mile long continuous corridor from I-94 to I-43 in Southeast Wisconsin, including 27 loop detector stations in the northbound and 26 in the southbound with a median gap of 2 miles.

• Identified 169 crashes resulted from lane changes by reviewing 2012 crash reports.

• Traffic data
  o Time: 5 -10 minutes before the crash
  o Variables: volume, speed and occupancy
  o Format: 1 minute traffic data
  o Location: next slide
One upstream station (A) and two downstream stations (B and C) nearest to the crash location.

Note: For the sake of continuity, the crash with any gap longer than 5 miles among Gap 1, 2 and 3 was removed.
Data Collection

- Four non-crash events were selected for each crash to serve as the controls.
- The traffic data at the same time, on the same day of the week, and at the same location as the crash were extracted.

<table>
<thead>
<tr>
<th></th>
<th>Station A</th>
<th>Station B</th>
<th>Station C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash Date-14 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crash Date-7 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crash Date</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crash Date+7 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crash Date+14 days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Processing

The mean, standard deviation and coefficient of variation of the 1 minute data were calculated for volume, speed and occupancy and were named with a four-part nomenclature method.

For example, ASAV stands for the average flow for the subject lane at the immediate upstream detector.
Data Processing

• It is anticipated that between-lane traffic differentials are critical to lane change decisions, and they were captured by dividing the average values of the subject lane by the average values of the target lane at the same detector location.

• Between-lane traffic differences are formulated by three letters with the first letter representing detector location, the second letter representing traffic variable, and the last letter “R” representing the ratio. For example, the ratio of flow at detector A, named AVR, is equal to ASAV/ATAV.

• In total, there were 63 variables (54 lane specific ones and 9 ratio ones)
Model Building Procedure

- A stratum is composed by a crash and its non-crash events, and this stratification feature makes it questionable to adopt popular variables selection methods (e.g. decision tree, random forest).

- To screen statistically significant variables and avoid the issue of collinearity, the univariable conditional logistic regression was applied for each of the 63 variables.

- Six variables significant at 10% level were selected.

- Among the six variables, two sets of relatively independent variables were obtained based on correlation test.

<table>
<thead>
<tr>
<th>Variable Set</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>ASSO, BSAV, CTSV, CTAS, CVR</td>
</tr>
<tr>
<td>#2</td>
<td>BTAV, CTSV, CTAS, CVR</td>
</tr>
</tbody>
</table>
• Stepwise selection method in the conditional logistic regression was conducted for both variable sets, and two models were developed.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable Set</th>
<th>Selected Variables</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>#1</td>
<td>BSAV, CTSV, CVR</td>
<td>168.432</td>
</tr>
<tr>
<td>2</td>
<td>#2</td>
<td>CTSV, CTAS, CVR</td>
<td>170.224</td>
</tr>
</tbody>
</table>

• The three variables contained in Model 2 represent the effect of speed, flow variation, and flow ratio that were more consistent with previous findings and the principles of traffic flow theory.
### Model results for Model 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Pr &gt; $\chi^2$</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAS</td>
<td>0.09011</td>
<td>0.0373</td>
<td>1.094</td>
</tr>
<tr>
<td>CTSV</td>
<td>0.20452</td>
<td>0.0118</td>
<td>1.227</td>
</tr>
<tr>
<td>CVR</td>
<td>1.20044</td>
<td>0.0217</td>
<td>3.322</td>
</tr>
</tbody>
</table>

- **CTSV**: standard deviation of volume in the target lane at detector C
- **CTAS**: average speed in the target lane at detector C
- **CVR**: flow ratio at detector C
- **Hazard Ratio**: it describes the ratio of change in the probability for a crash with one unit change in one variable. A hazard ratio greater than 1 means that the chance for a crash increases as the value of that variable increases.
1. All three variables are associated with traffic conditions from the furthest downstream location, Station C.

2. The contribution of the standard deviation of flow in the target lane at the furthest station to lane-changing related crashes may allude to more large gaps created by varying headways measured by a higher fluctuation of 1-min traffic volume.

3. The higher speed in the target lane contributes to such crashes.

4. Five out of six variables from two models are related to traffic flow, indicating the strong correlation between flow-related variables and lane-changing related crashes.
Questions?