Comparison Between Traditional Methods and the EB-SPF Method for Identifying High-Crash-Risk Intersections

In-Kyu Lim Ph.D., P.E.
CO-TED

2013 VASITE Annual Meeting
Outline

• Roadway Safety Management Process
• Overview of Network Screening Methods
• Research Motivation and Scopes
• Analysis and Results
• Conclusions
Roadway Safety Management Process

1. Network Screening
2. Diagnosis
3. Select Countermeasures
4. Safety Effectiveness Evaluation
5. Prioritize Projects
6. Economic Appraisal
Network Screening

• Purpose
  For reviewing a transportation network to identify and rank sites needed potential safety improvements

• Screening Methods
  ▪ Crash Frequency (CF)
  ▪ Crash Rate (CR)
  ▪ Rate-Quality Control (RQC)
  ▪ Equivalent Property Damage Only (EPDO)
  ▪ Empirical Bayes - Safety Performance Functions (EB-SPFs)

Traditional Methods
Screening Methods - Traditional

- Crash Frequency (CF)

\[ CF_i = \sum \# Crash \]

- **Strengths**

- **Weaknesses**

Does not account for traffic volume

Does not account for RTM bias
Regression-To-the-Mean (RTM)

Does this actual effectiveness due to the treatment? No!!

Perceived effectiveness of treatment
Regression-To-the-Mean (RTM)

Safety Improvement Project

- RTM Reduction
- Perceived effectiveness of treatment
- Actual Reduction due to Treatment

Observed Crash Frequency

Years
Screening Methods - Traditional

- Crash Rate (CR)

\[ CR_i = \frac{\sum \# Crash}{MEV_i} \]

\[ MEV_i = \frac{AADT}{1,000,000} \times (n) \times (365) \]

- Strengths

- Weaknesses
Screening Methods - Traditional

- Rate-Quality Control (RQC)

\[ CCR_i = ACR + z \sqrt{\frac{ACR}{MEV_i}} + \frac{1}{2 \cdot MEV_i} \]

\[ Critical \ Ratio_i = \frac{CR_i}{CCR_i} \]

- **Strengths**

- **Weaknesses**

\[ CR \]

\[ Avg. CR \]

\[ CCR \]
Screening Methods - Traditional

- **EPDO**
  
  \[ EPDO_i = \sum w_s \times \#Crash_{s,i} \]
  
  \( w_s \) = weight factor for severity level \( s \)

- **Strengths**

- **Weaknesses**

  - **Strengths**

  - **Weaknesses**

  
  **Sample EPDO Weights**

<table>
<thead>
<tr>
<th>Severity</th>
<th>Cost</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal (K)</td>
<td>$4,008,900</td>
<td>542</td>
</tr>
<tr>
<td>Injury (A/B/C)</td>
<td>$82,600</td>
<td>11</td>
</tr>
<tr>
<td>PDO (O)</td>
<td>$7,400</td>
<td>1</td>
</tr>
</tbody>
</table>

  **Example**

<table>
<thead>
<tr>
<th>Sites</th>
<th>Crash Severity</th>
<th># Crash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fatal</td>
<td>Injury</td>
</tr>
<tr>
<td>Site 1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Site 2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Site 3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
• Empirical Bayes – Safety Performance Functions (EB-SPFs)

1. Safety Performance Functions (SPFs)

“SPFs are regression equations that estimate crash frequency as a function of annual average daily traffic (AADT)”

\[ \text{Crash Frequency} = \exp(a + b \times \ln \text{AADT}) \]
Screening Methods – HSM

• Empirical Bayes – Safety Performance Functions (EB-SPFs)

2. Empirical Bayes

“EB is procedures for statistical inference in which the prior distribution is estimated from the data”

3. EB-SPFs

“EB-SPFs use a weight factor, which is a function of the SPF over-dispersion parameter, to combined the two estimates into a weighted average”

\[ N_{\text{expected}} = w \times N_{\text{SPF}} + (1 - w) \times N_{\text{observed}} \]

\[ w = \frac{1}{1 + k \times \left( \sum N_{\text{SPF}} \right)} \]

\[ k = \text{over-dispersion parameter from the associated SPF} \]
Screening Methods – HSM

• EB-SPFs

\[ N_{\text{expected}} = w \times N_{\text{SPF}} + (1 - w) \times N_{\text{observed}} \]

\[ w = \frac{1}{1 + k \times \left( \sum N_{\text{SPF}} \right)} \]

\[ k = \text{over-dispersion parameter from the associated SPF} \]

\[ \text{PSI}_i = N_{\text{expected},i} - N_{\text{SPF},i} \]

➢ Strengths

➢ Weaknesses

requires more data included as crash, demand, roadway inventory and traffic data.

requires SPF calibrated to local conditions.
How many State use the EB-SPF in U.S.?

- Only 2 – 3 State transportation agencies partially implemented,
- Over 90% of transportation agencies continuously use the traditional network screening methods

Why?
- Lack of data
- Limitation of data access, compile and/or manipulation
Research Motivation and Scope

• Motivation
  - Compare traditional methods for identifying high crash-risk locations against the EB-SPF Method
  - Propose most reliable traditional network screening methods as the results of the EB-SPF

• Scopes
  - Four traditional methods included (1) Crash Frequency, (2) Crash Rate (CR), (3) Rate-Quality Control and (4) Equivalent Property Damage Only (EPDO)
  - Analysis is limited for four-leg intersections categorized by area (urban/rural) and traffic control device (traffic signal/two-way stop)
Data and Virginia SPFs

• Data Source
  - VDOT Oracle dB, called HTRIS (Highway Traffic Records Information System)
  - Consists of 10 sub-systems included crash, roadway inventory and traffic

• Virginia SPFs
  - Since 2008, VDOT contributes efforts to develop Virginia SPFs with VCTIR

  VA SPF Development Status
  Two-lane highway - completed
  Intersection - completed
  Multilane highway - final review
  Freeway - final review

<table>
<thead>
<tr>
<th>Intersection Types</th>
<th>Safety Performance Functions</th>
<th>Dispersion (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban 4-Leg Traffic Signal</td>
<td>$C_{FP} = \exp \left( -7.6234 + 0.6746 \cdot \ln \text{AADT}<em>{\text{max}} + 0.3435 \cdot \ln \text{AADT}</em>{\text{norm}} \right)$</td>
<td>0.222</td>
</tr>
<tr>
<td>Rural 4-Leg Traffic Signal</td>
<td>$C_{FP} = \exp \left( -6.9588 + 0.6746 \cdot \ln \text{AADT}<em>{\text{max}} + 0.2530 \cdot \ln \text{AADT}</em>{\text{norm}} \right)$</td>
<td>0.217</td>
</tr>
<tr>
<td>Urban 4-Leg 2-Way Stop</td>
<td>$C_{FP} = \exp \left( -6.0723 + 0.4558 \cdot \ln \text{AADT}<em>{\text{max}} + 0.3470 \cdot \ln \text{AADT}</em>{\text{norm}} \right)$</td>
<td>0.428</td>
</tr>
<tr>
<td>Rural 4-Leg 2-Way Stop</td>
<td>$C_{FP} = \exp \left( -3.4940 + 0.3593 \cdot \ln \text{AADT}<em>{\text{max}} + 0.3935 \cdot \ln \text{AADT}</em>{\text{norm}} \right)$</td>
<td>0.293</td>
</tr>
</tbody>
</table>
In HTRIS, 1,670 intersections under 4 categories as below were extracted using SQL.

<table>
<thead>
<tr>
<th>Intersection Type</th>
<th>Number of Intersections</th>
<th>Variables</th>
<th>Min.</th>
<th>Max.</th>
<th>Average</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban 4-leg Traffic Signal</td>
<td>647</td>
<td>Total Crash</td>
<td>1</td>
<td>58</td>
<td>9.34</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Majr&lt;/sub&gt;</td>
<td>3,692</td>
<td>113,339</td>
<td>29,347.4</td>
<td>16,434.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Minn&lt;/sub&gt;</td>
<td>258</td>
<td>41,090</td>
<td>8,188.1</td>
<td>6,665.8</td>
</tr>
<tr>
<td>Rural 4-Leg Traffic Signal</td>
<td>153</td>
<td>Total Crash</td>
<td>1</td>
<td>17</td>
<td>3.87</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Majr&lt;/sub&gt;</td>
<td>1968</td>
<td>33,629</td>
<td>12,924.4</td>
<td>6,130.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Minn&lt;/sub&gt;</td>
<td>297</td>
<td>12,681</td>
<td>3,477.9</td>
<td>2,360.0</td>
</tr>
<tr>
<td>Urban 4-Leg 2-Way Stop</td>
<td>439</td>
<td>Total Crash</td>
<td>1</td>
<td>17</td>
<td>2.28</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Majr&lt;/sub&gt;</td>
<td>540</td>
<td>95,693</td>
<td>12,244.4</td>
<td>12,246.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Minn&lt;/sub&gt;</td>
<td>239</td>
<td>22,704</td>
<td>1,470.2</td>
<td>1,732.9</td>
</tr>
<tr>
<td>Rural 4-Leg 2-Way Stop</td>
<td>431</td>
<td>Total Crash</td>
<td>1</td>
<td>10</td>
<td>1.77</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Majr&lt;/sub&gt;</td>
<td>305</td>
<td>35,272</td>
<td>6,620.8</td>
<td>5,756.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AADT&lt;sub&gt;Minn&lt;/sub&gt;</td>
<td>216</td>
<td>7,201</td>
<td>923.0</td>
<td>652.1</td>
</tr>
</tbody>
</table>
Analysis and Results

- Distribution of 1,670 intersection PSIs using the EB-SPF
Analysis and Results (cont)

- Scatter Plots of output values
# Comparison Measures (1)

## Pearson Correlation Coefficient (r)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Crash Frequency</th>
<th>Crash Rate</th>
<th>EPDO</th>
<th>RQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB-SPF (All)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.494</td>
<td>.376</td>
<td>.252</td>
<td>.576</td>
</tr>
<tr>
<td>EB-SPF (PSI &gt; 0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.868</td>
<td>.153</td>
<td>.386</td>
<td>.713</td>
</tr>
</tbody>
</table>

<sup>a</sup> All: total intersections (n = 1,670).

<sup>b</sup> PSI > 0: positive PSI intersections (n = 842).
Comparison Measures (2)

2 Correct and false identification percentage of selecting top 1%, 5% and 10% hot-spots

<table>
<thead>
<tr>
<th>Top Unsafe Locations</th>
<th>Method</th>
<th>Correct Identification % (Count)</th>
<th>False Identification % (Count)</th>
<th>MAE Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crash frequency</td>
<td>76.5% (13)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td>1% (17 locations)</td>
<td>Crash rate</td>
<td>6.9% (1)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>52.9% (9)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>0.0% (0)</td>
<td>29.4% (5)</td>
<td></td>
</tr>
<tr>
<td>5% (84 locations)</td>
<td>Crash frequency</td>
<td>67.9% (57)</td>
<td>8.3% (7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash rate</td>
<td>20.2% (17)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>92.6% (78)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>67.9% (57)</td>
<td>9.5% (8)</td>
<td></td>
</tr>
<tr>
<td>10% (167 locations)</td>
<td>Crash frequency</td>
<td>65.9% (110)</td>
<td>15.6% (26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash rate</td>
<td>23.6% (40)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>75.4% (126)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>71.3% (119)</td>
<td>13.8% (23)</td>
<td></td>
</tr>
</tbody>
</table>

* Unsafe locations are those with positive PSI values.

* Correct identification when a location identified as being in top high crash-risk locations by a corresponding method is found in the top unsafe locations.

* False identification when a location is identified as being in top high crash-risk locations by a corresponding method is found in the safe locations (i.e., locations with PSI value ≤ 0).
Comparison Measures (3)

3  Rank based Mean Absolute Error (MAE)

\[
\text{MAE}(\text{Rank}) = \frac{1}{n} \sum_{i=1}^{n} |\text{Rank(EB - SPF)}_{i} - \text{Rank(T)}_{i}|
\]

- \text{MAE(\text{Rank})} = \text{mean absolute error in ranks}
- \text{Rank(EB-SPF)} = \text{rank of location I based on the PSI from the EB-SPF method}
- \text{Rank(T)} = \text{rank of location I based on performance measure from traditional methods}
- n = \text{number of locations varying depending on the specified top percent (1, 5, and 10%)}
Comparison Measures (3)

3 Rank based Mean Absolute Error (MAE)

Performance Comparisons of four Traditional Methods

<table>
<thead>
<tr>
<th>Top Unsafe Locations</th>
<th>Method</th>
<th>Correct Identification % (Count)</th>
<th>False Identification % (Count)</th>
<th>MAE Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% (17 locations)</td>
<td>Crash frequency</td>
<td>76.5% (13)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash rate</td>
<td>6.9% (1)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>52.9% (9)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>0.0% (0)</td>
<td>29.4% (5)</td>
<td></td>
</tr>
<tr>
<td>5% (84 locations)</td>
<td>Crash frequency</td>
<td>67.9% (57)</td>
<td>8.3% (7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash rate</td>
<td>20.2% (17)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>92.6% (78)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>67.9% (57)</td>
<td>9.5% (8)</td>
<td></td>
</tr>
<tr>
<td>10% (167 locations)</td>
<td>Crash frequency</td>
<td>65.9% (110)</td>
<td>15.6% (26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash rate</td>
<td>23.6% (40)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RQC</td>
<td>75.4% (126)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPDO</td>
<td>71.3% (119)</td>
<td>13.8% (23)</td>
<td></td>
</tr>
</tbody>
</table>

* Unsafe locations are those with positive PSI values.
* Correct Identification when a location identified as being in top high crash-risk locations by a corresponding method is found in the top unsafe locations.
* False Identification when a location is identified as being in top high crash-risk locations by a corresponding method is found in the safe locations (i.e., locations with PSI value ≤ 0).
Conclusions

• The Crash Rate (CR) method performed poorly in identifying the top 1, 5, and 10% of hot-spots. Thus, crash rate method does not recommend to identifying hot-spots.

• The Crash Frequency (CF) method performed the best in identifying the top 1% but false identification increased at the top 5 and 10%.

• The Rate-Quality Control (RQC) method performed the best in top 5 and 10% and no false identification in the all three levels of categories.
Conclusions

• The EB-SPF method recommends to use for identifying high crash intersections whenever it is feasible.
This research paper will be pressed on Transportation Research Record (TRR) in 2014.

Any Technical Assistant for the EB-SPF, please contact:

In-Kyu Lim, Ph.D., P.E.
Senior Highway Safety Engineer
CO-TED
e-mail: In-Kyu.Lim@vdot.virginia.gov

Young-Jun Kweon, Ph.D., P.E.
Research Scientist
VCTIR
e-mail: Young-Jun.Kweon@vdot.virginia.gov
Thank You