Analysis of Naturalistic Driving Study Data: Challenges and Opportunities

Paul P. Jovanis, Ph.D.
Professor
Civil and Environmental Engineering
Larson Institute
Pennsylvania State University
Many Collaborators

* Dr. Kun-Feng Wu – National Research Council Fellow, Turner-Fairbank Highway Research Laboratory
* Dr. Jonathan Aguero - Escuela de Ingeniería Civil Universidad de Costa Rica
* Dr. Frank Gross – VHB Inc.
* Adam Greenstein – Maryland State Highway Administration
* Prof. Venky Shankar – Penn State, *Special Thank You*
What is a Naturalistic Driving Study?  
On-board Devices, Normal Driving, and Roadway Data
Past and On-going Naturalistic Driving Study Data Collection

* USA
* Europe
* Canada
* Japan
* Australia
* China

* Passenger cars/light trucks
* SHRP 2 NDS
* Safety Pilot (UMTRI; PC and Truck)
* 100-car Study
* FOTs 2000 and earlier (UMTRI)
* Volvo (EU)
* Teen Drivers
* Older Drivers
* Large trucks (VTTI)
* Motorcycles (VTTI)
* Bicycles (Australia proposal)
Basic Idea

Many people believe NDS data make safety analysis easier. . . . Because detailed information is available about the actions of the instrumented vehicle and driver
Basic Idea is a Fallacy

Many people believe NDS data make safety analysis easier... Because detailed information is available about the actions of one involved driver and vehicle.

But the opposite is true...

1. Data not collected to test hypotheses
2. Safety analysis structure changes
3. Many new paradigms possible
1. Develop Safety Performance Functions (SPFs) – relate traffic level to expected crashes per year
2. Network Screening – identification of sites with promise
3. Site Diagnosis – identify factors contributing to crashes
4. Countermeasure selection and effectiveness evaluation
Analogy

Comparing 1960’s “4-step travel forecasting process” with contemporary micro-simulation and activity-based approaches

Some elements of problem are similar, but wider and deeper questions can be asked and addressed
Seminar Goal

Introduce the Naturalistic Driving Study technique to you, providing examples of new research questions and paradigms that are evolving as a result of the technique.
Basic Idea

Many people believe NDS data make safety analysis easier. . .

Because detailed information is available about the actions of one involved driver and vehicle:

But the opposite is true . . .

1. Data not collected to test hypotheses

2. Safety analysis structure changes – need to supplement crash events with additional events - SURROGATES

3. Many new paradigms possible - initial studies of driver-based crash modeling
What are Crash Surrogates?

Normal Driving/Baseline

Safety-Relevant Events

Crash Surrogates

Crashes
The Relationship Between Crashes and Crash Surrogates

The most well-known crash surrogate – conflicts at intersections.

- **General Criterion:** “an situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movement remain unchanged.”

- **Specific Condition:** Time-to-collision less than 1.5 seconds.

\[ \lambda = \pi \times c, \quad \pi = \lambda / c, \quad \pi: \text{conversion factor} \]

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>Car-Car</th>
<th>Car-Unprotected road user</th>
</tr>
</thead>
<tbody>
<tr>
<td>1+2</td>
<td>3.2\times10^{-5}</td>
<td>15.3\times10^{-5}</td>
</tr>
<tr>
<td>3+4</td>
<td>11.1\times10^{-5}</td>
<td>3.2\times10^{-5}</td>
</tr>
</tbody>
</table>

Source: Hyden (1987)
Study Goal: Crash Surrogates

1. **Identify** *well-defined crash surrogates*, which have common etiologies to crashes (but not the same crash outcome).

2. **Better understand crash causation** by studying well-defined crash surrogates and crashes together at detailed level.
What are Crash Surrogates? – Crash Generating Process

Identifying and Screening Crash Surrogates from NDS Data

1. How to search for safety-related and surrogate events in a large NDS dataset?
   * Labor intensive for reviewing video data.

2. Develop screening strategies
   * A need to facilitate comparison across studies.
   * Self-induced sample selection bias.

3. Which safety-related events should we consider to be surrogates?
Study Design: 63 Short Trips with Roadway Departure Safety-related Events

Event 37

Before

False Alarms

During Event

After

True Event of Interest
First Screening – The Concept

Normal Driving vs. Events of Interest

- Grey: Normal Driving
- Red: Threshold = 5
- Green: Events of Interest

Risk Index vs. Frequency

Risk Index: 0 to 50
Frequency: 0 to 6000

Nominal Driving Threshold = 5
Events of Interest
## Identifying Event of Interest (1):
The Trade-off between Sensitivity and Specificity

<table>
<thead>
<tr>
<th>Event of Interest</th>
<th>Normal Driving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test as Event of Interest</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Test as Normal Driving</td>
<td>Missed</td>
</tr>
</tbody>
</table>

\[
\text{Sensitivity} = \frac{\text{Total Events of Interest}}{\text{Missed}}
\]

\[
\text{Missed} = \frac{\text{Total Events of Interest}}{\text{Total Events of Interest}}
\]
Identifying Event of Interest (2): The Trade-off between Sensitivity and Specificity

<table>
<thead>
<tr>
<th></th>
<th>Event of Interest</th>
<th>Normal Driving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test as Event of Interest</td>
<td>Sensitivity</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Test as Normal Driving</td>
<td>Miss it</td>
<td>Specificity</td>
</tr>
</tbody>
</table>

Specificity = \[
\frac{\text{Total Normal Driving Events}}{\text{Total Normal Driving Events}}
\]

False Alarm = \[
\frac{\text{Total Normal Driving Events}}{\text{Total Normal Driving Events}}
\]
Self-induced Sample Selection Bias

What if young drivers depress brake pedal harder or swerve sharper than other drivers?


Young Drivers

Events of Interest

Normal Driving Events

Screening Measure

Others

kernel = epanechnikov, bandwidth = 0.2899

Events of Interest

Normal Driving Events
Application Using NDS Data from NHTSA Web Site

* Developed data set using kinematic variables for 63 road departure events from VTTI 100 car study obtained from NHTSA web site

* Combined with data on verified events from SHRP 2 S-01 project (video verified by VTTI)

* Applied algorithm in series of papers
Separate Events of Interest from Normal Driving

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Maximum Lateral Acceleration (LATM)</th>
<th>Maximum Difference in Lateral Acceleration (LATD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>&gt;= 0.0g</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>&gt;= 0.1g</td>
<td>100.00%</td>
<td>4.80%</td>
</tr>
<tr>
<td>&gt;= 0.2g</td>
<td>98.41%</td>
<td>22.40%</td>
</tr>
<tr>
<td>&gt;= 0.3g</td>
<td>92.06%</td>
<td>33.60%</td>
</tr>
<tr>
<td>&gt;= 0.4g</td>
<td>71.43%</td>
<td>60.80%</td>
</tr>
<tr>
<td>&gt;= 0.5g</td>
<td>46.03%</td>
<td>77.60%</td>
</tr>
<tr>
<td>&gt;= 0.6g</td>
<td>38.10%</td>
<td>84.00%</td>
</tr>
<tr>
<td>&gt;= 0.7g</td>
<td>26.98%</td>
<td>89.60%</td>
</tr>
<tr>
<td>&gt;= 0.8g</td>
<td>12.70%</td>
<td>95.20%</td>
</tr>
<tr>
<td>&gt;= 0.9g</td>
<td>7.94%</td>
<td>96.80%</td>
</tr>
<tr>
<td>&gt;= 1.0g</td>
<td>4.76%</td>
<td>98.40%</td>
</tr>
<tr>
<td>&gt;1.0g</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Using Different Screening Measures

- Lat30D ROC area: 0.7512
- Lat01M ROC area: 0.8953
- Lat30M ROC area: 0.7096
- Lat10D ROC area: 0.948
- Lat10M ROC area: 0.8573
- Yaw30D ROC area: 0.6549

Reference
The effect of lateral acceleration on crash risk is different.
1. Look for variables that are influential to crash risk over time.

2. Select thresholds based on sensitivity and specificity.

<table>
<thead>
<tr>
<th></th>
<th>Crashes</th>
<th>Events of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test as Crashes</td>
<td>Sensitivity</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Test as Events of Interest</td>
<td>Miss it</td>
<td>Specificity</td>
</tr>
</tbody>
</table>
Second Screening Results
Verification: Theoretical Analysis Structure

First Screening

Normal Driving

Safety-Relevant Events

Classification; Second Screening; Verification;

Surrogate Event, $Y_1 = 1$

Crash, $Y_2 = 1$

Near Crash

$$\Pr(Y_2 = 1 | Y_1 = 1, X) = 1$$

$$\text{Prob}(Y_2 = 1 | Y_1 = 1, X) = 1$$

$$\text{Cov}(Y_1, Y_2 | X_1, X_2) = 1$$

$$\text{Prob}(Y_2 = 1 | Y_1 = 1, X) >> 0$$

$$\text{Cov}(Y_1, Y_2 | X_1, X_2) >> 0$$
Identifying Surrogate Events Quantitatively

Case 32 After First Screening

Measurement Duration (Every 100 units = 10 seconds)

- Original Trip
- Lat10D > 0.4g during event
- Lat01M > 0.3g during event
- Lat10M > 0.3g during event
- Lat30M > 0.4g during event
- Yaw30D > 4 degree during event
- Lat30D > 0.4g during event

Northwestern University  May 16, 2013
Conditional Crash Probability


\[
\text{Surrogate Event, } Y_1 = 1 \quad \xrightarrow{\text{Crash}} \quad \text{Crash, } Y_2 = 1 \\
\quad \xrightarrow{\text{Near Crash}} \quad \text{Near Crash, } Y_2 = 0
\]
The Conversion Factor Table for Single Vehicle Run-off-road Events

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Daytime/Nighttime</th>
<th>Maximum Lateral Acceleration Difference During 3-second Window</th>
<th>Conversion Factors</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Night</td>
<td>0.9 - 0.8g</td>
<td>0.08</td>
<td>2</td>
<td>0.16</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Night</td>
<td>1.0 - 0.9g</td>
<td>0.13</td>
<td>1</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Day</td>
<td>1.0 - 0.9g</td>
<td>0.09</td>
<td>5</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Night</td>
<td>&gt; 1.0g</td>
<td>0.57</td>
<td>3</td>
<td>1.71</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Day</td>
<td>&gt; 1.0g</td>
<td>0.56</td>
<td>8</td>
<td>4.48</td>
<td>4</td>
</tr>
</tbody>
</table>
# Another Approach: Driver-based Analyses

<table>
<thead>
<tr>
<th>Outcome (0/1)</th>
<th>Length</th>
<th>Time</th>
<th>Event Attributes (as many as needed)</th>
<th>Context Variables (as many as needed)</th>
<th>Driver Attributes (as many as needed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Another Approach: Driver-based Analyses

Context variables: road functional class, ramp, urban/rural, day/night;

Event variables: wet/dry; distraction; fatigue

Driver variables: gender, education, years of driving experience
Another Approach: Driver-based Analyses

<table>
<thead>
<tr>
<th># of Outcomes (count)</th>
<th>Total Length (veh. mi.)</th>
<th>Total Time (Veh.hours)</th>
<th>Event Attributes (as many as needed)</th>
<th>Context Variables (as many as needed)</th>
<th>Driver Attributes (as many as needed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bottom Line: the analyses I have shown you look nothing like the traditional road safety management analyses!
Naturalistic driving studies are promising:

* Surrogate analyses critical; require knowledge of vehicle kinematics
* Include driver attributes in screening criteria
* Driver-based models offer additional detail not currently available
* New skills needed to fully utilize information contained in data
* When data from SHRP 2 become fully available, many exciting studies are possible
THANK YOU!

QUESTIONS??
References


