Teen Drivers: Spatial Pattern of Crashes

Hemin J. Mohammed, Ph.D.
Road Safety Resource Coordinator
University of Kansas Transportation Center

2019 Mid-Continent Transportation Research Symposium
Ames, Iowa
Road traffic injuries are a major threat, especially for teens aged 15–19. Traffic crashes involving teen drivers have been a worldwide concern for several decades.

Traffic crashes in the US are a leading cause of death for teens since 2001. In 2015, 2,333 teens were killed (six teens every day). In 2016 this increased by 3.6%.
In Kansas, KDOT’s Strategic Highway Safety Plan (SHSP) 2015, identified teen drivers as one of its emphasis areas to reduce crashes.

The number of crashes involving teen drivers in Kansas increased by 644 (10,715 to 11,356) from 2013 to 2016.

The SHSP goals were not met.
Curry et al. (2011)
• 822 teen drivers were involved in 795 serious crashes
• Driver error was by far the most common reason for crashes (95.6%)
• Among crashes with a driver error, a teen made the error (79.3%) of the time

Singh (2018) categorized these errors into:
• Recognition errors
• Decision errors
• Performance errors
• Non-performance errors
The distinguishing characteristics of teen drivers associated with high-rate crashes are:

- Inexperience
- High risk-taking behaviors
Teen drivers have:

- Higher hazard response times
- Greater speed and steering variability
- Higher crash rates in inclement weather, at horizontal curves, and at intersections
• Modeling spatial characteristics at:
  – County level
  – State level
DATA

- The Fatality Analysis Reporting System (FARS) database
- Kansas Department of Transportation (KDOT)
- U.S. Census Bureau Database
- Kansas State Department of Education
Population and Licensed Teenagers by Age and Gender
## Crashes by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Crashes</th>
<th>Crashes Involving Teen Drivers</th>
<th>Crashes Involving Teen Drivers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>60,562</td>
<td>11,765</td>
<td>19.43</td>
</tr>
<tr>
<td>2011</td>
<td>60,270</td>
<td>10,978</td>
<td>18.21</td>
</tr>
<tr>
<td>2012</td>
<td>58,373</td>
<td>10,587</td>
<td>18.14</td>
</tr>
<tr>
<td>2013</td>
<td>59,265</td>
<td>10,472</td>
<td>17.67</td>
</tr>
<tr>
<td>2014</td>
<td>59,938</td>
<td>10,508</td>
<td>17.53</td>
</tr>
<tr>
<td>2015</td>
<td>61,440</td>
<td>10,709</td>
<td>17.43</td>
</tr>
<tr>
<td>2016</td>
<td>62,390</td>
<td>11,172</td>
<td>17.91</td>
</tr>
<tr>
<td>Total</td>
<td>422,238</td>
<td>76,191</td>
<td>18.04</td>
</tr>
</tbody>
</table>
Counties & Crashes
Modeling Spatial Relationships

- Ordinary Least Squares (OLS)
- Geographically Weighted Regression (GWR)
POTENTIAL EXPLORATORY VARIABLES

Potential Exploratory Variables

- DVMT on all types of roads
- DVMT on rural non-state roads
- Household income
- Miles of all types of roads
- Miles of rural non-state roads
- Number of families (income is below poverty level)
- No. of High Schools
- No. of postsecondary schools
- No. of workers >15 years & commute to work
POTENTIAL EXPLORATORY VARIABLES

• No. of non-commercial trucks
• No. of passenger cars
• Population of >15 years in the labor force
• Population of 18-24 yrs old < high school degrees
• Population of counties
• Population of Females >15 years in the labor force
• Population of Males >15 years in the labor force
• Population of teens
• Precipitation
EXPLORATORY REGRESSION

Exploratory Variables

- X1
- X2
- X3
- X18

Test all variable combinations for:

- Redundancy
- Significance
- Bias
- Performance

Outputs
EXPLORATORY REGRESSION

Every variable should be statistically significant
to be included in the model

Six variables were statistically significant

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>The average number of passenger cars</td>
</tr>
<tr>
<td>DVMT_ALL</td>
<td>Average DVMT on all types of roads</td>
</tr>
<tr>
<td>RD</td>
<td>Miles of rural non-state roads in counties</td>
</tr>
<tr>
<td>HIGH.SCHOOL</td>
<td>Number of High Schools</td>
</tr>
<tr>
<td>POP:TEEN</td>
<td>The population of teens</td>
</tr>
<tr>
<td>POP</td>
<td>The population of counties</td>
</tr>
</tbody>
</table>
# EXPLORATORY REGRESSION

<table>
<thead>
<tr>
<th>Statistical Measure</th>
<th>Formula/Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Inflation Factor (VIF)</td>
<td>Indicates the degree to which a predictor variable inflates the standard errors of other predictors.</td>
</tr>
<tr>
<td>Adjusted R Squared</td>
<td>Adjusted to account for the number of predictors in the model.</td>
</tr>
<tr>
<td>Joint F-Statistic p-value</td>
<td>Measures the overall significance of the predictors.</td>
</tr>
<tr>
<td>Joint Wald Statistic p-value</td>
<td>Similar to the F-Statistic but for Wald statistics.</td>
</tr>
<tr>
<td>Koenker (BP) Statistic p-value</td>
<td>Measures the overall significance of the predictors.</td>
</tr>
<tr>
<td>Jarque-Bera p-value</td>
<td>Tests for normality of the residuals.</td>
</tr>
</tbody>
</table>
### Ordinary Least Squares Regression Model

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R-Squared</td>
<td>0.9101</td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.9084</td>
</tr>
<tr>
<td>Joint F-Statistic</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Joint Wald Statistic</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Koenker (BP) Statistic</td>
<td>0.0133</td>
</tr>
<tr>
<td>Jarque-Bera Statistic</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept</th>
<th>RD</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>-1.0655</td>
<td>0.0192</td>
<td>1.8058</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.05693</td>
<td>0.00713</td>
<td>0.06028</td>
</tr>
<tr>
<td>Probability</td>
<td>&lt;0.0001*</td>
<td>0.00826*</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Robust SE</td>
<td>0.07802</td>
<td>0.00717</td>
<td>0.06540</td>
</tr>
<tr>
<td>Robust Pr.</td>
<td>&lt;0.0001*</td>
<td>0.00868*</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>VIF</td>
<td>--------</td>
<td>1.08688</td>
<td>1.08688</td>
</tr>
</tbody>
</table>

VIF: Variance Inflation Factor
ORDINARY LEAST SQUARES REGRESSION (OLS) MODEL

- Chase County underpredicted
- Comanche, Stanton, Decatur overpredicted

\[
CRASH = e^{\left(-1.0655 + 0.0192 \left( \sqrt{\frac{RD}{\ln RD}} \right) + 1.8058 \left( \log (\ln PC) \right) \right)}
\]
GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) MODEL

\[ E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon \]

\[ E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon \]
GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) MODEL

Coefficient of average number of passenger cars
GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) MODEL

Coefficient of miles of rural non-state roads
VALIDATION OF MODELS

• Validate by number of crashes in 2017
• Use number of registered passenger cars in each county in 2017
• Use miles of non-state roads in each county in 2017
• OLS underestimated overall number of crashes was by 3.66 %
• GWR underestimated overall number of crashes was by 2.94 %
VALIDATION OF MODELS

- Rawlins County (5 crashes in 2017)
- Predicted number (7.65 crashes)
- Overestimated by 53% (2.65 crash)

<table>
<thead>
<tr>
<th>Percentage</th>
<th>OLS Underestimated</th>
<th>OLS Overestimated</th>
<th>GWR Underestimated</th>
<th>GWR Overestimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1%</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>(1-4.9)%</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>(5-9.9)%</td>
<td>16</td>
<td>9</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>(10-24.9)%</td>
<td>18</td>
<td>21</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>(25-49.9)%</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>&gt; 50%</td>
<td>4</td>
<td>10</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>54</td>
<td>51</td>
<td>54</td>
</tr>
</tbody>
</table>

Among 30 counties that had the highest number of crashes, only Jefferson and Wyandotte had the predicted number of crashes off by >25%
PREDICTION BY MODELS

• Predict number of crashes in 2026
• Use growth factor
• Assume no other changes
• OLS predicts 3.11% reduction
• GWR predicts 3.37% reduction
FINDINGS AND RECOMMENDATIONS

• Developed statistically significant models (OLS and GWR) to predict future teen-related crashes at the county and state levels

• Among 18 related exploratory variables, the number of miles of non-state roads and the number of passenger cars were found to be statistically significant

• OLS and GWR models predicted a three percent reduction in the number of crashes, statewide by 2026

• This research provides a useful indicator for related parties to identify where they can target their resources
FUTURE STUDIES

• This research methodology can be used to analyze other subsets such as alcohol-related crashes, and older driver crashes

• Conducting temporal analysis using ArcGIS to compare crash rates between teen drivers and other age groups

• Comparing the findings of the current research with the network screening results