A statistical look at gender and age differences as related to the injury crash type on low-volume roads

F. Russo, S. A. Biancardo, M. Busiello, M. De Luca & G. Dell’Acqua
Dipartimento di Ingegneria Civile, Edile e Ambientale, University of Napoli Federico II, Italy

Abstract

The research presented here is addressed to develop only one safety performance function (SPF) from the perspective of driver gender for three identified main crash types (head-on/side collisions, rear-end collisions, single-vehicle run-off-road crashes) that is able to predict the injury crash rate on low-volume roads. According to the crash police reports, it came in sight that males and females differ in terms of their psychological attributes and, consequently, their response to the crash risk can change producing different effects on the severity.

The analysis was divided into two phases: the first deals with SPF calibration, while the second concerns SPF validation. A total length of 355 km was used in the first phase involving 5 years of the crash database (2003–2007), to a total of 95 injury crashes which led to 136 injuries (63% male only drivers, 8% female only drivers and 29% female+male drivers) and 9 deaths (78% male only drivers and 22% female+male drivers). A total length of 295 km was used in the second phase involving 3 years of the crash database (2008–2010), to a total of 73 injury crashes which led to 120 injuries (68% male only drivers, 4% female only drivers and 28% female+male drivers) and 4 deaths (75% male only drivers and 25% female+male drivers). GEE was adopted to calibrate SPF. Mean width, mean speed at each analyzed road segment, and a numerical variable “SLEH” reflecting the identified road “Surface” (dry/wet), “Light” conditions (day/night), geometric “Element” (tangent segment/circular curve) and “Human” factors (gender/age/number drivers) all together when the crash happened, were introduced in the predictive safety model looking toward gender and age drivers.

Keywords: female and male drivers, response to crash risk, low-volume roads.
1 Introduction

Road safety continues to be a prickly problem for transport system despite many years of initiatives to reduce the crashes number.

Munala and Maina [1] described that the economic cost of road crashes and injuries is estimated to be at 1% of the gross national product in low-income countries, 1.5% in middle-income countries and 2% in high-income countries. Recommended international practice stipulated that drivers should take 20–30-min breaks after every 1.5–2 h of driving.

Low-volume roads (LVRs) as presented in this paper are vital part of an infrastructure system: these roads must be well designed, well-constructed, and properly maintained to create minimal adverse effects, to be cost-effective in the long term, and to have acceptable maintenance and repair costs. Limited resources often exist for LVRs, and frequent questions as whether to use paved or gravel surfaces are formulated [2]. Some procedures were developed for several years now as cost-effective alternatives for evaluating road surface conditions: i.e. porous asphalt was suggested as a feasible solution during critical weather conditions (rain, ice, etc.) to offer better friction in wet conditions, reduction of splash, spray and aquaplaning in overtaking, reduction of flash on wet surfaces, and better sound absorption [3].

Many researchers have dealt with driver speed behavior on rural roads to identify all the possible factors that may affect safety conditions during travel. These factors can be directly linked to personal choices, vehicle conditions, and infrastructure/environmental features [4]. Driving is a complex activity involving not only cognitive processes but also emotional, and social interaction process.

Nordfjærn et al. [5] investigated differences in psychological safety factors such as attitudes and behaviour among professional and non-professional drivers using questionnaire survey distributed by mail to a randomly selected sample. Arnau-Sabatés et al. [6] revealed the existence of some risk factors for road drivers such as speed, external circumstances, distraction, and alcohol. More males than females had a higher tendency to take risks in all the exposed risky factors.

Das et al. [7] established that the use of alcohol but also drugs can impair judgment and increase the possibility of other high-risk behaviors such as speeding, risk-taking, and violating traffic rules and thereby contribute to involvement in Road traffic crashes.

Scott-Parker et al. [8] explored the mileage (duration and distance), car ownership, and punishment behaviors of young licensed provisional drivers and their relationships to risky driving, crashes, and offenses.

Several studies have focused on the issue of gender and its relation to crashes. In a study by Al-Balbissi [9], male-to-female crash ratios were calculated and the results indicated considerably high male rates (3.42 times more than female rates).

In this paper relationships between the influence of gender and the infrastructure/environmental circumstances on the crash dynamics were found to predict the crash phenomenon. Based on the research results, the driver is the
most important active safety system, its optimal psycho-physical conditions are among the best guarantees in order to avoid accidents and a critical situation.

The risk associated with how specific road/environmental/human factors can influence on the consequences of an investigated crash type on the studied road segment was studied. Only one Safety Performance Function was performed to predict the injury crash rate by varying gender/age/number of drivers, crash type, “Light” (day/night), road “Surface” (dry/wet) conditions, and road element.

2 Literature review

Eustace and Wei [10] compared the age and gender of drivers involved in fatal crashes and the corresponding driving errors that contributed to these crashes. The youngest drivers tended to be carrying the highest number of passengers when they were involved in fatal crashes. Failing to stay in the proper lane and driving too fast were the two most frequent driver operating errors contributing to fatal crashes for male and female drivers.

Kirk and Stamatiadis [11] conducted a research to identify the specific driving maneuvers in specific types of crashes involving younger drivers. They explained that drivers’ inexperience is the largest single contributor to their increased crash rates.

Chandraratna and Stamatiadis [12] performed an analysis to evaluate potential problem maneuvers that may lead to higher crash involvement. Older drivers are more likely to be involved in crashes related with left turn maneuvers and gap acceptance compared with younger drivers: older male drivers are safer than older female with a side passenger.

Stigson [13] presented differences in average crash severity, distribution of crash severity, and injury outcomes, based on an independent safety rating of roads, also taking road type and speed limit into consideration. Furthermore, he evaluated differences in injury risk, based on the distribution of crash severity.

Quddus [14] contributed to the debate on the relationship between average speeds, speed variations, and accident rates. Speed variation was found to be statistically and positively associated with accident rates. A 1% increase in speed variation was associated with a 0.3% increase in accident rates, ceteris paribus.

Mohammadi [15] studied the influence of age, belt use, time of day, and type of vehicle road. Analysis indicated that the female drivers have a much better road safety record than male drivers, especially with regards to their involvement in severe traffic accidents. Female drivers were found to be generally safer drivers than their male counterparts; male drivers who did not use seat belts had a higher involvement rate in road traffic crashes.

Mccartt et al. [16] showed that age and experience have important and independent effects on crash risk. The studies consistently found that teenage drivers had dramatically higher crash rates than older drivers.

The purpose of this experimental analysis was intended to recognize that on existing roadways a potential combination of the values of geometric, kinematic, traffic features and human factors (gender/age/number drivers) can contribute to the occurrence of a crash. The helpfulness of these variables in reducing the
number of crashes was examined through an experimental analysis using statistical approach. The calibration of only one injury crash rate prediction model for specific target collision type by varying human factors is important because of the range of harms that are caused by different collision types and different combination of gender/age/number drivers.

Potential dangerous road scenarios can be identified for male or female only drivers. The number of possible strategies for road safety improvements depend on the specific available resources at the time and they can be equal to the variables employed in the model on which it can actually work: before-after evaluations [17–20] allow us to check the safety benefits of improvements.

3 Data analysis

Since 2003 the authors carry out a large crash data collection on 650 km of two-lane rural roads in Southern Italy located on flat area with a vertical grade of less than 6%. Table 1 shows synthetic descriptive statistics of the main features observed on the investigated road network where L is the total element length, CCRm is the curvature change rate of a homogeneous segment [21], CCRs is the curvature change rate of the circular curve, W is the travel lanes plus shoulders, R is curve radius.

Table 1: Descriptive statistics of the mean features observed and measured.

<table>
<thead>
<tr>
<th></th>
<th>L [km]</th>
<th>CCRm [gon/km]</th>
<th>CCRs [gon/km]</th>
<th>W [m]</th>
<th>R [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min Value</td>
<td>0.3</td>
<td>9.60</td>
<td>0.004</td>
<td>5.23</td>
<td>15</td>
</tr>
<tr>
<td>Max Value</td>
<td>2.03</td>
<td>662.66</td>
<td>4246.7</td>
<td>9.95</td>
<td>5000</td>
</tr>
</tbody>
</table>

After a thorough analysis of crash data, focusing in particular on the relationship between the injury crash rate/driver gender scenario, the injury crash database was divided into subsystems (see second column of Table 2) and in further substrates (see third column of Table 3) to maximize the reliability of the SPFs and their statistical parameters during the calibration and validation phase reflecting and predicting real phenomenon.

A total length of the analyzed network equal to 355 km was used in the calibration phase involving 5 years of the crash database (2003–2007), to a total of 95 injury crashes which led to 9 deaths (78% male only drivers and 22% female+male drivers) and 136 injuries (63% male only drivers, 8% female only drivers and 29% female+male drivers). In particular:

- 60% involved crashes by male only drivers with
  - 85 injuries of which 38% involved head-on/side collisions, 62% single-vehicle run-off road crashes, and
  - 7 deaths of which 14% involved head-on/side and rear-end collisions, 72% single-vehicle run-off road crashes,
- 9% for female only drivers with
  - 11 injuries of which 18% involved head-on/side collisions, 82% single-vehicle run-off road crashes;
31% for female+male drivers (84% one man and one woman drivers and 16% two men and one woman drivers) with
- 40 injuries of which 67% involved head-on/side collisions, 3% rear-end collisions, 30% single-vehicle run-off road crashes, and
- 2 death involved single-vehicle run-off road crashes.

The category “crashes with male and female drivers” refers to three identified conditions: a) one man and one woman drivers; b) two women and one man drivers; c) two men and one woman drivers.

A total length of the analyzed network equal to 295 km was used in this validation phase involving 3 years of the crash database (2008–2010), to a total of 73 injury crashes, which led to 4 deaths (75% male only drivers and 25% female+male drivers) and 120 injuries (68% male only drivers, 4% female only drivers and 28% female+male drivers). In particular:

- 68% involved crashes by male only drivers with
  - 82 injuries of which 82% involved head-on/side collisions, 5% rear-end collisions, 13% single-vehicle run-off road crashes, and
  - 3 deaths of which 100% involved head-on/side collisions;
- 5% for female only drivers with
  - 4 injuries of which 25% involved head-on/side collisions, 75% single-vehicle run-off road crashes;
- 27% for female+male drivers (75% for one man and one woman drivers, 19% two men and one woman drivers and 6% two women and one man drivers) with
  - 34 injuries of which 94% involved head-on/side collisions, 6% single-vehicle run-off road crashes, and
  - 1 death involved single-vehicle run-off road crashes.

The highest proportion of crashes during the eight-study period involved cars (75%), of which 56% were driven by only men, 7% by only female and 37% by female+male (44% male and 56% female). Two wheeled vehicles were involved in 14% of cases, of which 75% were driven by only men, 3% by only women and 22% by female+male drivers (80% male and 20% female). The remaining 11% of all accidents involved trucks, tractors, buses, and agricultural vehicles, of which 65% were driven by only men and 35% by female+male (100% male).

Figures 1–3 show the corresponding percentage of only men drivers, only female drivers and female+men drivers involved in crashes by varying the crash type, the vehicle type and the injuries/deaths number.

To reflect specific combinations of road geometric-function-environmental circumstances and human factors (gender/age/number drivers) which influence the consequences of a crash type, the authors designated an original numerical variable, “SLEH” to denote various crash settings. This crash scenario variable (SLEH) was used to calibrate only one safety predictive model reflecting the identified road “Surface” (dry/wet), “Light” conditions (day/night), geometric “Element” (tangent segment/circular curve) and “Human” factors (gender/age/number drivers) when the crash happened, as provided by related police reports. The Safety Performance function predicts by varying the gender
drivers, crash type and road scenario the injury crash rate (ICR) for homogeneous segments equal to the number of injury crashes per year per $10^8$ vehicles/km (crash frequency over traffic exposure). During the study period the mean crash rate value makes the maximum on the circular curves during nighttime on wet road surface conditions when a head-on/side injury collision happens.

Before moving to the calibration phase, a technique to filter anomalous injury crash rates for each subsystem was adopted, calculated using the Vivatrat method [22] widely used in geotechnical engineering.

![Crash profiles for only male drivers.](image)

![Crash profiles for only female drivers.](image)

![Crash profiles for male+female drivers.](image)
The method [22] is based on estimates of ranges of values that can be considered fluctuations of the “regular” measures compared with values estimated as “abnormal” for each homogeneous substrate within each homogeneous subsystem. The subsystem reflects the crash type and in the case of study three main crash types were identified on the investigated road network: head-on/side collisions, single-vehicle run-off-road crashes, rear-end collisions. 3 subsystems for the only male drivers sample, for the only female drivers sample, and for the female+male drivers sample were defined as well as identified in Table 2.

Because of overdispersion of crash data, a generalized estimating equation and an additional log linkage equation were adopted to calibrate SPF. It was opted to perform one injury crash prediction model to better reflect the influence of gender and infrastructure/environmental circumstances on the crash dynamics for low-volume roads as observed on the investigated road network, well functional and well-designed to maximize the statistical reliability and effectiveness of the prediction models during the validation phase, as follows:

The values associated with the SLEH variable, presented as one of the explanatory variables of the safety performance function, permits interested parties to examine the risk associated with how specific road/environmental/human factors can influence on the consequences of an investigated crash type on the studied road segment. The SLEH code is higher for the more dangerous investigated substrates in terms of collected injury and death, and lower for those less dangerous. The values of the “SLEH” variable (see last column of Table 2) were set for each substrate as follows:

- counts of the substrates (road scenarios) number (n_j) (27 substrates were defined)
- calculating for each substrate, the mean injury crash rate (ICR_j) over the period of the study
- identification on whole database of substrates, the highest mean injury crash rate (ICR_{max,j})
- SLEH code equal to 10 for the substrate with maximum ICR_{j,max} and for all those remaining, a value SLEH_j (Eq. (1)) is as follows:

$$Code \ SLEH = 10 \cdot \frac{ICR_j}{ICR_{j,\text{max}}},$$

AIC was used a measure of the goodness of fit of a statistical model shown in Table 3 and it describes the tradeoff between the accuracy and the complexity of a model. It includes a penalty that discourages over fitting; thus, the best models are those with lowest AIC value together with the smallest amount of the explanatory variables. The safety equations in Table 3 that predicts the number of injury crashes per year per 108 vehicles/km (crash frequency over traffic exposure) by varying gender drivers and crash type on two-lane rural roads located on the flat area, vertical grade of less than 6% and without transition curves between tangent segments and circular curves shows good AIC and Pearson dispersion values.
Table 2: SLEH Code according to gender, crash type and scenario.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Subsystem (Crash Type)</th>
<th>Substrate (road scenario)</th>
<th>Mean value of crash rate</th>
<th>Code SLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury crashes with only male drivers</td>
<td>head-on/side collisions</td>
<td>WNC</td>
<td>47.45</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WDC</td>
<td>30.77</td>
<td>3.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDC</td>
<td>25.03</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNC</td>
<td>22.25</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>21.25</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDT</td>
<td>21.01</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>single-vehicle run-off-road crashes</td>
<td>WNC</td>
<td>45.75</td>
<td>4.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WDC</td>
<td>36.57</td>
<td>3.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNC</td>
<td>30.90</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDC</td>
<td>16.69</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>10.36</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>rear-end collisions</td>
<td>WDC</td>
<td>40.94</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>13.96</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDT</td>
<td>11.13</td>
<td>1.17</td>
</tr>
<tr>
<td>Injury crashes with only female drivers</td>
<td>head-on/side collisions</td>
<td>WNC</td>
<td>30.90</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>17.24</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>single-vehicle run-off-road crashes</td>
<td>WNT</td>
<td>65.36</td>
<td>6.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>39.57</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDT</td>
<td>8.84</td>
<td>0.93</td>
</tr>
<tr>
<td>Injury crashes with male + female drivers</td>
<td>head-on/side collisions</td>
<td>WNC</td>
<td>94.97</td>
<td>10.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WDC</td>
<td>82.73</td>
<td>8.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDC</td>
<td>58.05</td>
<td>6.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WDT</td>
<td>48.64</td>
<td>5.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>19.91</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDT</td>
<td>7.79</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>single-vehicle run-off-road crashes</td>
<td>WDT</td>
<td>25.58</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNT</td>
<td>17.69</td>
<td>1.86</td>
</tr>
</tbody>
</table>

Symbols: Wet road surface + Night + Curve (WNC), Wet road surface + Daylight + Curve (WDC), Dry road surface + Daylight + Curve (DDC), Dry road surface + Night + Curve (DNC), Dry road surface + Night + Tangent segment (DNT), Dry road surface + Daylight + Tangent segment (DDT), Dry road surface + Daylight + Curve (DDC), Wet road surface + Night + Tangent segment (WNT), Wet road surface + Daylight + Tangent segment (WDT).

Each explicative variable inserted into the models has a different effect on predicting the injury crash rate, as may be observed from the value of the
coefficient and the algebraic sign. The only variable always negatively correlated to the $Y$ dependent variable is the mean width of the carriageway instead of SLEH variable and the mean speed on the homogeneous segment.

Table 3: SPF from driver gender perspective by varying the crash type.

<table>
<thead>
<tr>
<th>SPF</th>
<th>AIC</th>
<th>Max log-likelihood</th>
<th>Deviance</th>
<th>Pearson Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = e^{(0.28 \cdot SLEH - 0.25 \cdot MW + 0.05 \cdot MS)}$ (2)</td>
<td>7.11</td>
<td>-195.22</td>
<td>0.04</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

SLEH

“SLEH Code” in Table 2

MW

mean width of the travel lanes+shoulders in meters

MS

mean speed on the analyzed road segment in km/h

4 Results

A total length of the analyzed network equal to 295 km was used in this validation phase involving 3 years of the crash database (2008–2010). The validation procedure consisted of the follows steps:

a) assessment of the residual value at each homogeneous road segment as the difference between the predicted ICR using Eq. (2) and the real ICR;

b) analysis of the std.dev. ($\sigma$) and mean ($\mu$) of the sample values of ICR using the “3$\sigma$” method (see Figure 4). The distribution of residuals confirms the percentages of the crash frequency distribution came within the “action limits” ($\mu \pm 3\sigma$) and “attention limits” ($\mu \pm 2\sigma$);

c) diagrams of performance (see Figure 5) where on the x-axis there is the observed ICRs and on the y-axis there is the predicted ICRs by Eq. (2);

d) assessment of the mean residual of ICR prediction model as the mean of differences ($D_i$) between observed and predicted ICRs at each homogeneous road segment;

e) assessment of the mean absolute deviation (MAD), which is a constant value equal to the sum of the absolute values $D_i$ divided by the number of road segments;

Figure 4: Control chart of the residuals for the injury crash rates.

Figure 5: Performance diagram of Eq. (2).
f) assessment of the mean squared error (MSE), which is a constant value equal to the sum of $D_i^2$ values divided by the number of road segments. It was noted that more than 90% of the residuals of the crash rate prediction model fall within the range $\mu \pm 2\sigma$, and that all the values fall within the range $\mu \pm 3\sigma$. In particular, it was noted how 75% of residuals fall within the range $\mu \pm \sigma$ and the maximum residual is equal to 76.51. This condition is confirmed by the performance diagrams in the Figure 4. The reliability of the regression equations is also completed by the mean error value less than 6.00 ICR and the low value obtained for $MAD$ (7.22) and $MSE$ indicators (8.50). It can be concluded that safety performance function is statistically significant.

5 Conclusions

Potential countermeasures that can help to reduce the frequency of the most critical type of crash on the investigated road network were suggested. The case shown below refers to the most dangerous studied scenario. Circular curves during nighttime on wet road surface conditions when a head-on/side injury collision happens. In the diagram, the $y$-axis shows the predicted ICR, while the $x$-axis shows an independent variable of the predictive model equal to mean speed on the geometric element. Fig. 6 at first presents three series of curved lines with a constant width of 8m for only male drivers (blue line), only female drivers (pink line), male+female drivers (orange line).

![Figure 6: Abacus to predict ICR by varying speed and gender drivers.](image)

Moving along the same line to the left, there can be a reduction in crash density on the roadway segment with no change in roadway width but only in speed; moving downward along a vertical line, we can observe different behaviors by varying gender drivers. In addition to lower the injury crash rate we suggested to modify WNC road scenario in DDT scenario (see results in Table 4 and Figure 6) by varying SLEH Code as shown in Table 2, with no change in width and speed.
Injury crash rate prediction models that involve the total number of crashes during an examined period are useful to help agencies responsible for establishing funds for infrastructure improvements, for prioritizing project decisions based on issues such as greatest return on investment and safety.

Table 4: Changing road scenario.

<table>
<thead>
<tr>
<th>Gender Drivers</th>
<th>Most dangerous scenario WNC</th>
<th>Scenario DDT</th>
<th>D ICR\textsubscript{m} From WNC 80 Km/h To WNC 50 Km/h</th>
<th>D ICR\textsubscript{m} From WNC 80 Km/h To DDT 50 Km/h</th>
<th>D ICR\textsubscript{m} From DDT 80 Km/h To DDT 50 Km/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>only male drivers</td>
<td>Code SLEH V = 80 Km/h Predicted ICR\textsubscript{m}</td>
<td>Code SLEH V = 80 Km/h Predicted ICR\textsubscript{m}</td>
<td>Code SLEH V = 80 Km/h Predicted ICR\textsubscript{m}</td>
<td>Code SLEH V = 80 Km/h Predicted ICR\textsubscript{m}</td>
<td>Code SLEH V = 80 Km/h Predicted ICR\textsubscript{m}</td>
</tr>
<tr>
<td>only female drivers</td>
<td>3.25</td>
<td>35.63</td>
<td>7.42</td>
<td>2.21</td>
<td>16.40</td>
</tr>
<tr>
<td>male + female drivers</td>
<td>5</td>
<td>21.90</td>
<td>4.56</td>
<td>0.93</td>
<td>11.49</td>
</tr>
<tr>
<td>Reference Code</td>
<td>SLEH</td>
<td>Predicted ICR\textsubscript{m} V = 80 Km/h</td>
<td>Predicted ICR\textsubscript{m} V = 80 Km/h</td>
<td>Predicted ICR\textsubscript{m} V = 80 Km/h</td>
<td>Predicted ICR\textsubscript{m} V = 80 Km/h</td>
</tr>
</tbody>
</table>

References


