Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 27 (2017) 1088–1096

Transportation Research Procedia

www.elsevier.com/locate/procedia

20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary

Estimating an Injury Crash Rate Prediction Model based on severity levels evaluation: the case study of single-vehicle run-off-road crashes on rural context

Francesca Russo[a], Roberta Di Pace[b*], Gianluca Dell'Acqua[a], Stefano de Luca[b]

[a]*Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, via Cluadio 21 Napoli, Italy*
[b]*Department of Civil Engineering, University of Salerno via Giovanni Paolo II, 132 Fisciano (SA) Italy*

Abstract

In general in case of crash situations the quality of collected data is very limited and several information are usually unreliable. Thus it is recognised that a significant effort is required in order to improve the quality of the crash prediction models moreover a crucial role is played by the identification of the factors influencing the crashes occurrence and the levels of severity estimation. In this paper two injury crash rate prediction models related to single-vehicle run-off-road crashes type are calibrated and in particular significant attributes estimated are identified not only with roadway geometric characteristics and surface conditions, but also with gender/number-of-drivers. To this aim a survey of injury crashes on two-lane rural roads collected in the Southern Italy was considered and analysed. Finally before the calibration step, a preliminary analysis of the data was provided through the estimation of the levels of severity by multinomial logit; in fact by this model only segments with highest values of severity are identified and involved in the calibration procedure.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: Injury Crash Rate Prediction Model; Multinomial Logit; Survey data; Rural Roads Network; Single-Vehicle Run-Off-Road Crashes

* Corresponding author. Tel.: +39 089 963420; fax: +39 089 968748
E-mail address: rdipace@unisa.it

2352-1465
Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.
10.1016/j.trpro.2017.12.159

1. Introduction an motivation

Reducing crashes on highways has always been one of the most important concerns for transportation engineers during the processes of planning, design, construction, and maintenance. Providing a safe driving environment is indeed not only a responsibility, but also one of the top priorities for all highway projects (Chuo et al. 2009). Many authors dealt with specific crash types to identify causes, effects and appropriate countermeasures. Moreover collected data of crashes are very limiting in terms of information thus a significant effort to improve our ability in understanding the factors influencing the crashes occurrence is required.

Srinivasan *et al.* (2009) showed, for example, that approximately 27% of fatal crashes on the U.S. highway system occurred along horizontal curves on two-lane rural highways. Approximately 70% of curve-related fatal crashes were single-vehicle crashes in which the vehicle left the roadway and struck a fixed object or overturned, and 11% of curve-related fatal crashes were head-on crashes.

Hu and Yang (2011) have studied the Golumd–Kunlun Mountain section of the low-volume Qinghai–Tibet highway. Their analysis showed that 72.5% of crashes occurred on straight sections and 9.2% on curves. Vehicle rollover occurred in 47.5% of all accidents, mostly on a combination of vertical and horizontal curves and a small-radius curve.

The objective of the study presented here is to develop an injury crash rate prediction model for the only single-vehicle run-off-road crash type, as a first approach, that was observed on the studied two-lane rural roads in the Salerno Province. Two injury crash rate prediction models were calibrated for maximizing the reliability and effectiveness of the results reflecting proper combinations of road geometric-function-environmental circumstances and human factors (gender/number drivers) which influenced the consequences of a crash type. First of all, each rural road belonging to the study network was divided into homogeneous segments according to the a constant curvature change rate definition; each road contained from one to three road homogeneous segments.

A first analysis of the crash data was performed through the application of a multinomial logit formulation (Mc Fadden, 1974). In particular three (discrete) levels of severity were considered: (a) less than 3 injuries, (b) more than 2 and less than 4 injuries and finally (c) more than 4 injuries and fatality. Thus a statistical model aiming to derive the probability of crash can be considered (Cafiso and D'Agostino, 2016) . The main assumption is that the probability of occurrence of the crash is higher than the probability of occurrence of all other events (the set is composed by all possible levels of severity). In order to estimate these probabilities a function for the probability distribution must be specified; in particular the function is characterised by an alternative specific constant, a vector of estimable coefficients and the error term accounting for unobservable factors. Depending on the considered distributional form of the error term a specific model may be derived. Considering this distribution as normal could be a most common assumption however in this case the resulting model will be a probit model which is, as well known, a model very complicate to be estimated (see Ben-Akiva and Lerman, 1985) then most frequently the error term is considered being generalised extreme value (GEV) distributed. This kind of model may be solved in a closed form through maximum likelihood methods; the resulting model is the multinomial logit model (McFadden, 1974). This considered model is characterised by the irrelevant alternatives (IIA) assumption (Train, 2003) which implies that unobserved factors are uncorrelated over the alternatives. Even though it may be argued that this assumption is very limiting in the meantime it makes the multinomial logit model particularly convenient. Based on all previous considerations then this model formulation was applied leading for the prediction of the levels of severity of each segment within each rural road in the considered network. The significant result refers to homogeneous road segments with highest severe degree further investigated and considered in the Injury Crash Rate Prediction (*ICR*) Model calibration.

Furthermore, next step focused on the investigation of the more severe road scenarios, as explained before, to define and plan road operations for improving safety conditions: before and after safety evaluations can be used to check the safety benefits of improvements carried out on the roadways, within budget constraints for improvement or safety compliance investments for future operation. The study aims to integrate, innovate and bring new results in addition to those already obtained in a previous experimental analysis (Dell'Acqua, 2011, 2015; Russo et al. 2014, Russo et al. 2016).

EI-Basyouny and Sayed (2006) compared two types of regression techniques for predicting crashes: the traditional negative binomial (TNB) and the modified negative binomial (MNB). While the TNB approach assumes that the shape parameter of the negative binomial distribution is fixed for all locations, the MNB approach assumes that the shape parameter can vary from one location to another.

Qin (2012) applied an alternative crash modeling approach: quantile regression (QR) in the context of a count data model. QR model for crash count data confirmed that crash predictors have varying impacts on the different areas of the crash distribution and the marginal effects of covariates provide a more direct observation of changes in the quantity. Kweon and Oh (2011) developed a modeling approach to identify promising road segments for safety improvement through speed management strategies and to illustrate how to select segments on the basis of model results. The study involved the application of four statistical techniques (generalized additive model, negative binomial model, linear model, and empirical Bayes method) in three sequential steps to data collected on a 190-km section of expressway in South Korea. There is a large body of research that analyzes the effect of passenger age and gender on young driver fatal crash risk (Hough et al. 2008), the main differences between older female and male drivers (Rosenbloom 2006), the weight of age and experience of drivers in driving mileage (McCartt et al. 2009). Islam et al. (2006) explored, for example, the differences in the severity of driver-injury between male and female drivers, across three different age groups for single-vehicle accidents involving passenger cars.

The reminder of the paper is organised as follows: data collection and the whole preliminary analyses are described in section 2; in section 3 is discussed in more detail the methodological framework whilst section 4 focuses on the numerical results; finally, section 5 discusses conclusions and future perspectives.

2. Data collection and preliminary analyses

The crash data used in this research for calibrating ICR Prediction model for the single-vehicle crashes type involved almost 2000 km of two-lane rural roads in Southern Italy located in the flat area with a vertical grade of less than 6%. The model helps to predict the number of the crashes for year for km for 10^8 vehicles (crash frequency over traffic exposure). Homogeneous road segments were defined by a constant curvature change rate (CCR in gon/km) defined as the sum of the absolute values of angular changes in horizontal alignment divided by total length of road section. Curvature Indicator *CI* was associated at each homogeneous road segment reflecting CCR values; CI is a measurement of the curvature change rate of the homogeneous road segment and it was associated a value between 1 and 3: 1 for low road horizontal curvature (CCR < 50gon/km); between 1 and 2 for medium road horizontal curvature (50gon/km ≤ CCR ≤ 300gon/km; between 2 and 3 for high road horizontal curvature (CCR > 300gon/km).

The road segments involved in the experimental study have a mean value of the Annual Average Daily Traffic of 4,000 vpd/h, length of 3.5 km, mean speed of 65 km/h, mean road width (travel lanes plus shoulders) of 7.00 m with a minimum of 5.00 m and a maximum of 12.00 m, and a mean value for the curve radius of 200 m. Crash data, refer to a 5 years study period (2006-2010), and they were made available by the Administration of the Province of Salerno. Over the total length of the network analyzed, 998 injury crashes can be observed from 2006 to 2010, with 1,819 injuries and 71 deaths, and a mean injury crash rate of 40 (the number of injury crashes per year per km per 10^8 vehicles on the horizontal homogeneous segment of two-lane rural roads). Three main crash types was identified as a whole: frontal/side-head-on collisions, single-vehicle crashes (vehicle exits the roadway and either strikes a fixed object or overturns), and rear-end collisions. Analysis showed that frontal/side-head-on collisions occurred in 52%, while single-vehicle run-off-road crashes occurred in 33%. Rear-end collisions made up only 15% of cases. In particular, focusing on "single-vehicle crashes" type as a first approach in the analysis, it was observed by analyzing the crash reports during the study period that 67% of crashes happened with male only drivers, 15% with female only drivers, and 18% with male and female drivers. In particular, the category "crashes with male and female drivers" refers to three identified conditions: a) one male and one female drivers; b) two females and one male drivers; c) two men and one female drivers.

Table 1 shows an overview of the analyzed injury crashes varying the crash's context (*scenario*) during the total study period, incorporating gender into the analysis, for which is shown the corresponding percentage of injury crashes (*n*), and relative minimum, mean, and maximum value of injury crash rate (*ICR*) during the 5-year period.

According to the results in Table 1, it can be observed that the maximum value for injury crash rate using gender analysis during the specific period is when only men drivers are involved in crashes on *wet road surface + daylight + curve* (max crash rate of 99.52). The minimum value for injury crash rate when female only drivers are involved in crashes for the relevant period is recorded on *dry road surface + daylight + tangent* segment scenario (crash rate of 0.53), while when female and male drivers are involved the scenario is with *wet road surface + night + curve* scenario (crash rate of 0.97). The maximum value of the crash rate when female only drivers are involved in a crash, as when female and male drivers are involved, is recorded for *dry road surface + night + tangent segment* scenario.

Table 1. Injury Crash Rates' summary using Scenario, Gender analysis for single-vehicle run-off-road crashes

Road scenario	n., %	Min ICR	Mean ICR	Max ICR
CRASHES INVOLVING NO MORE THAN THREE VEHICLES WITH ONLY MALE DRIVERS				
Dry road surface + Daylight + Curve (*DDC*)	15	0.58	10.97	22.77
Dry road surface + Daylight + Tangent segment (*DDT*)	16	0.58	5.55	10.11
Dry road surface + Night + Curve (*DNC*)	22	0.42	24.99	26.90
Dry road surface + Night + Tangent segment (*DNT*)	16	0.89	25.01	46.69
Wet road surface + Daylight + Curve (*WDC*)	9	2.05	34.88	99.52
Wet road surface + Daylight + Tangent segment (*WDT*)	4	1.25	11.00	33.94
Wet road surface + Night + Curve (*WNC*)	11	0.83	24.24	33.92
Wet road surface + Night + Tangent segment (*WNT*)	7	1.24	9.16	19.34
CRASHES INVOLVING NO MORE THAN THREE VEHICLES WITH ONLY FEMALE DRIVERS				
Dry road surface + Daylight + Curve (*DDC*)				
Dry road surface + Daylight + Tangent segment (*DDT*)	32	0.53	10.95	15.01
Dry road surface + Night + Curve (*DNC*)	13	2.52	3.20	6.90
Dry road surface + Night + Tangent segment (*DNT*)	13	3.97	34.90	33.92
Wet road surface + Daylight + Curve (*WDC*)	18	2.29	7.13	12.23
Wet road surface + Daylight + Tangent segment (*WDT*)				
Wet road surface + Night + Curve (*WNC*)	13	2.18	5.39	13.54
Wet road surface + Night + Tangent segment (*WNT*)	11	1.75	10.91	7.91
CRASHES INVOLVING NO MORE THAN THREE VEHICLES WITH FEMALE+MALE DRIVERS				
Dry road surface + Daylight + Curve (*DDC*)				
Dry road surface + Daylight + Tangent segment (*DDT*)				
Dry road surface + Night + Curve (*DNC*)	17	3.10	6.57	16.90
Dry road surface + Night + Tangent segment (*DNT*)	20	0.98	9.12	44.33
Wet road surface + Daylight + Curve (*WDC*)	35	1.47	9.26	30.90
Wet road surface + Daylight + Tangent segment (*WDT*)	11	3.05	11.47	25.58
Wet road surface + Night + Curve (*WNC*)	17	0.97	10.18	13.75
Wet road surface + Night + Tangent segment (*WNT*)				

A careful analysis of the database has shown that a wide variety of factors appear to influence or be associated with the crash dynamic. It was determined by deep examination and several statistical studies that "surface" (dry/wet) and "light" conditions (day/night), location of the detected crashes (tangent/circular curve element), lane width, horizontal curvature indicator (measurement of the curvature change rate), mean speed, gender factors and number of vehicles involved are significant and consistent variables to explain the effects of a crash.

Moreover, according to the procedure preliminarily described in the introduction based on the application of multinomial logit for the level of severity prediction, at each homogenous road segment was associated the occurrence probability of three severe events in terms of the injuries number: (a) less than 3 injuries, (b) more than 2 and less than 4 injuries and finally (c) more than 4 injuries and deaths. Therefore, only homogeneous road segments with highest severe degree (3) at each investigated road were selected and involved in the calibration phase of ICR prediction model. To maximize the reliability and effectiveness of the results in terms of Akaike information criterion (AIC) and Bayesian information criterion (BIC), two ICR prediction models for single-vehicle run-off-road crashes were calibrated: a) it was built one equation for wet road surface conditions and b) it was built one ICR prediction model for dry road surface conditions.

With reference to the multinomial logit the estimation results are displayed in Table 2. In particular it must be observed that as expected significant attributes may be distinguished in seasons, light condition, surface conditions, road characteristics (i.e curve; width, tortuosity in terms of curvature indicator; vertical grade). To be more precise constant parameters play a role only in event two and three; road vertical grade (negative), tortuosity (positive) and roadway width (positive) are significant only in alternative one and two however roadway width shows the same magnitude whereas other attributes are higher magnitude in alternative two; the speed (positive) is significant only in alternative three whereas the segment length is significant in both alternative one and two (negative as expected) and values in terms of magnitude are quite similar; finally in terms of road condition and surface condition curve is significant (positive) in alternative three whilst dry road surface is significant only in alternative two. Other relevant attributes refer to the light condition (night) which is significant in both alternative two and three (positive) with higher magnitude in alternative three. Moreover also seasons are positively significant: winter in first and second events with similar values, winter in all three alternatives with higher value in alternative two and similar lower values in alternative three. Finally, in the model are also embedded attributes related to the drivers characteristics such as the age which is positive and significant only in alternative two and the gender (male) which is positive only in alternative three. In terms of validation test all attributes are significant and the significance is shown in the brackets (t-test results); furthermore a more specific evaluation of the percent right was computed then a satisfying value of the 68% of percent right is obtained.

Table 2. Estimation results of the multinomial logit

Attributes	*First event*	*Second event*	*Third event*
Autumn Season	+ 1.44 *(+1.21)*	+ 1.62 *(+ 1.14)*	
Winter Season	+ 0.775 *(+ 0.90)*	+ 1.19 *(+ 1.03)*	+ 0.78 *(+ 1.01)*
Night (Light condition)		+ 0.43 *(+ 0.81)*	+ 3.29 *(+ 1.85)*
Curve			+ 0.66 *(+ 0.98)*
Dry road surface		+ 1.39 *(+ 2.42)*	
Driver's age		+ 5.7 *(+0.93)*	
Driver's gender (male)			+ 4.72 *(+ 31.05)*
Segment length	-0.13 *(-2.29)*	-0.17 *(-2.01)*	
Speed			+ 0.07 *(+ 0.63)*
Roadway width	+ 0.85 *(+ 1.14)*	+ 0.65 *(+ 2.00)*	
Curvature indicator	+ 0.56 *(+ 0.69)*	+ 1.51 *(+ 1.56)*	
Vertical grade	-0.34 *(-0.66)*	-1.18 *(-1.65)*	
Constant parameter		-3.52 *(-8.92)*	-7.91 *(-2.52)*
Rho-square	*0.690*		
Adjusted rho-square:	*0.610*		

3. Modelling framework

Because of crash data over-dispersion, a generalized estimating equation GEE with an additional log linkage equation was adopted to calibrate the ICR prediction model by involving only road homogeneous segments associated with high severe degrees of the occurrence probability of the event 3 as explained in the previous section.

GEE form is like GLM, but full specification of the joint distribution are not required, and thus no likelihood function. GEE form is as follows in Equation 1.

$$g(\mu_i) = x_i^T \cdot \beta \tag{1}$$

where

$g(\mu_i)$ is the link function defined as identity, log, logit, etc.

x_i represents a set of explanatory variables which can be discrete, continuous, or a combination

In particular by using GEE, a) correlated data are modeled using the same link function and linear predictor setup as in the case of independent responses, b)random component is described by the same variance functions as in the independence case; c) covariance structure of the correlated responses must also be specified and modeled.

Some GEE assumptions are as follows:
- The responses are $Y_1, Y_2, ..., Y_n$ are correlated or clustered;
- Covariates can be the power terms or some other nonlinear transformations of the original independent variables, can have interaction terms;
- The homogeneity of variance does not need to be satisfied;
- Errors are correlated;
- It uses quasi-likelihood estimation rather than maximum likelihood estimation (MLE) or ordinary least squares(OLS) to estimate the parameters, but at times these will coincide.

The research illustrated here follows a "network" approach for the safety analysis of the investigated road segments: a) identification of the injury critic crash type on the studied road network with highest frequency of occurrence on the roadway segments during the analysis period; b) identification of "black" roadway segments for a specific crash type, by using a injury crash rate prediction model, where the crash injury rate is higher than on the rest of the roads; c) injury crash rate prediction model.

Before moving to the calibration phase, a technique to filter anomalous injury crash rates for each of two subsystems of the entire dataset that returned the best results validation (I. ICRs refer to wet road surface; II. ICRs refer to dry road surface) was adopted. The 3σ method was used to remove anomalous injury crash rates and it makes it possible to check the homogeneity distribution around the mean, and the maximum deviation of frequency distribution is 3σ. Analyzing anomalous points it makes it possible to identify data far away mean distribution that should be investigated separately and by using different and proper procedure. As part of the literature review, a number of statistical anomaly detection approaches were developed to identify anomalous data before moving to the model calibration step for maximizing the goodness of fit including statistical control chart techniques (3-sigma outlier, moving range, SPI/CPI outlier), Grubbs, Rosner, and Dixon tests, and Tukey box plots. The ICRs values falling outside $mean \pm 3 \cdot std.dev.$ ($\mu \pm 3 \cdot \sigma$) are removed from two subsystems before moving to the calibration step. The results of the 3σ method showed that 3 values were rejected for ICRs concerning "dry road surface" scenario because greater than μ+3σ (equals 30), and 1 value was rejected for ICRs concerning "wet road surface" scenario because greater than μ+3σ (equals 50).

A numerical code "*GRCS*" was performed to reflect the global mean severity of the injury crash rate at each substrates of the subsystem defined that reflect the seriousness of each road scenario where injury crashes happened. The *GRCS* code was assessed by analyzing the average effects on the crash outcomes related to the *gender* factors (consequences on the severity of the crashes by varying number of drivers and gender) and *road crash scenario* (dry/wet conditions of the road surface + lighting conditions + road geometric site where the crashes were happened + curvature indicator CI).

The *GRCS* code is shown in Table 3 and it is one of the explanatory variables of the ICR prediction model, and it is higher for the high injury crash rates falling within the same subsystem, and lower for those with lower injury crash rates. The values of the "*SLEH*" were set for each substrate as follows:

- calculating for each substrate *i-th* of the subsystem *j-th* the mean injury crash rate ($ICR_{i,j}$) over the period of the study
- order each of twelve substrates, belonging to each of two subsystems, in increasing the mean value of injury crash rate ($ICR_{i,j}$);
- identification for each subsystem *j-th*, the highest mean injury crash rate ($ICR_{max,j}$);
- *GRCS code* assignment for each *i-th* substrate in the *j-th* subsystem with code= n_j for the *i-th* substrate in the *j-th* subsystem with highest mean injury crash rate, and for all those remaining, a value C_{GRCS_i} = Eq. (2) is as follows:

$$C_{GRCS\_i} = n_i \cdot \left(\frac{ICR_{i,j}}{ICR_{\max,j}} \right) \quad (2)$$

Table 3. GRCS code determinations according to gender drivers and road crash scenarios

			only male drivers	only female drivers	male+female drivers
Subsystem	Substrates	CI range	*Code GRCS*	*Code GRCS*	*Code GRCS*
Wet road surface	Daylight + Curve (*WDC*)	1	12	4.32	5.80
		1<CI<2	10.05	4.90	6.00
		2 < CI ≤ 3	8.01	5.35	6.12
	Daylight + Tangent segment (*WDT*)	1	5.90		7.90
		1<CI<2	5.01		6.80
		2 < CI ≤ 3	4.23		5.30
	Night + Curve (*WNC*)	1	10	3.52	5.55
		1<CI<2	8.2	4.00	4.20
		2 < CI ≤ 3	7.52	4.19	3.20
	Night + Tangent segment (*WNT*)	1	3.80	4.91	
		1<CI<2	3.10	4.20	
		2 < CI ≤ 3	2.70	3.78	
Dry road surface	Daylight + Curve (*DDC*)	1	5.20		
		1<CI<2	4.89		
		2 < CI ≤ 3	4.38		
	Daylight + Tangent segment (*DDT*)	1	2.01	4.82	
		1<CI<2	1.87	4.25	
		2 < CI ≤ 3	1.74	3.96	
	Night + Curve (*DNC*)	1	9.56	1.56	2.64
		1<CI<2	7.58	1.13	2.53
		2 < CI ≤ 3	6.82	1.03	2.01
	Night + Tangent segment (*DNT*)	1	10.89	12	5.16
		1<CI<2	7.86	11.75	4.78
		2 < CI ≤ 3	7.04	10.79	4.03

Akaike information criterion (AIC) and Bayesian information criterion (BIC) were mainly used to check the reliability of the model among many equations come out from several iterations. The preferred model was the one with the minimum AIC value. Hence AIC not only rewards goodness of fit, but also it includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over fitting; generally, model selection performance improved as sample sizes increased. The ability of AIC to select a true model rapidly increases with the sample size and with the decrease of the standard deviation of the sample size. Under unstable conditions such as small sample and large noise levels AIC outperforms BIC. The choice of the model that best fits the data is based on the AIC available values' comparison in small/medium samples as well as in the present study since the crashes are rare events, whilst on the BIC values in larger samples with high noise.

In addition to these criteria, other goodness-of-fit measures for GEE models were used: *p*-value to test significance of the regression coefficients (only coefficients with $p < 0.05$ were kept in the models), deviance, Pearson dispersion, Wald and Pearson χ^2 test.

4. Numerical results

GEE was adopted to calibrate ICR prediction models (see Equations 3 and 4) for each of two subsystems shown in the Table 3 for maximizing the effectiveness of the predicted values, reducing the residuals between observed and predicted ICR determinations (number of injury crashes per year per 10^8vehicles/km on the homogeneous road segment).

The AIC value for Equation 3 was 6.10, and BIC was 256.30. The goodness-of-fit measures of the final model were deviance 0.02 and Pearson dispersion 0.08,while AIC value for Equation 4 was 11.32, and BIC was 105.22.

The goodness-of-fit measures of the final model were deviance 0.04 and Pearson dispersion 0.00054. It should be remembered that Equs.2 and 3 do not apply to areas near the intersections that require ad hoc models.

$$ICR_{dry\_road\_surface} = e^{(-0.31 \cdot MW + 0.29 \cdot GRCS + 0.52 \cdot MS)} \quad (3)$$

$$ICR_{wet\_road\_surface} = e^{(-0.10 \cdot MW + 0.12 \cdot GRCS + 0.09 \cdot MS)} \quad (4)$$

where

- *MW* is the mean width of the travel lanes plus shoulders, in m;
- MS is the mean speed at each analyzed homogeneous roadway segment, in km/h;
- GRCS is available in the Table 2 and it is a factor that reflects human factors and crash road scenario.

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 *ICR* dependent variable is the mean width of the travel lanes. The equations 3 and 4 are to apply independently one by one on all investigated road homogeneous segments. Subsequently, different solutions can be considered: a) to sum all results derived from the application of the performed ICR prediction models for different infrastructure and environmental conditions for studying the global safety state on the whole road network; b) to study the crash injury rates for specific crash types under definite "Light" (day/night) and road "Surface" (dry/wet) conditions, gender/number of drivers, mean speed and roadway width, curvature indicator, by using definite equations 3 and 4. Equations 3 and 4 are an essential tool in assessing the best solution of road safety. The number of possible strategies is equal to the number of potential combination of the explanatory variables. Countermeasures for reducing injury crash rates can be suggested including awareness campaigns and real road structural operations. Awareness campaigns include direct actions on the population, such as training, pamphlets, and safety events tailored to the specific driver safety issues identified by age and gender. Specific awareness campaigns for male drivers, for example, can be implemented if critical geometric, environmental, functional circumstances are reflected in a higher GRCS code as shown in Table 3. For a critical scenario that involves females only (see Table 3), then specific strategies can be defined; such as training or posters targeted which emphasize the factors on the road that influence crashes and encourage a safe, prudent and attentive driving style. Disciplinary actions for unsafe driving can vary depending on whether a certain critical crash scenario is most frequently caused by male or female only drivers. Safe driving campaigns can be tailored to the specific needs of target gender depending on worst infrastructure features (road surface/light) that can cause high ICR and on which it should be worked since they have a greater impact for emphasizing hazard road factors during the campaign.

5. Conclusions and future perspectives

The objectives of this research are to calibrate two Injury crash rate prediction models to calculate the expected injury crash frequency over traffic exposure (injury crash rate) using a gender/number-of-drivers analysis of those involved in injury crashes on two-lane rural roads in the Southern Italy, infrastructure features and crash type analysis. Acceptable countermeasures can be suggested for road safety targets as structural operations and actions on the community with targeted awareness campaigns for safer driving that highlight major risk factors by gender and age. The research study confirmed that a frequent injury crash that may be due to road geometric failures is also strongly related to human factors such as degree of comfort, knowledge of the environment, and a driver's ability to perceive the coming road alignment. Knowing the critical crash (injury crashes with highest frequency of occurrence on a homogeneous segment during the analysis period) it is possible to identify accurate and precise countermeasures for those crash dynamics. So there would seem to be no need to sum all the crash types on the studied homogeneous segment during an examined period (frontal/side-head-on collisions or single-vehicle run-off-road crash or rear-end collisions) as the crashes are stochastically independent events: it should be enough to identify the critical injury crash type with the highest crash frequency on the investigated road network and to study an accurate safety performance function that reflects all the features when the crash happened in order to identify detailed countermeasures.

Future developments will be carried out for improving the effectiveness of injury crash rates prediction model as follows: a) assessment of the difference between the after and before proportions of the injury crash rates at each treatment site for a specific target collision type; b) assessment of the average difference between after and before proportions over all *n* treatment sites; c) assessment of the statistical significance of the average shift in proportion of the target collision type; d) calibrating injury crash rates prediction models for others crash types; e) validating procedures. The validation procedure focuses on residuals analysis; the residual is the difference between observed and predicted ICR to be assessed on rural road network out of the calibration step but reflecting the features of the first road network used. Several parameter evaluations can be considered as follows for analyzing the residuals values that's an essential tool to identify where the predictive models may miss the mark, over- or underestimating the real predicted ICR: a. CUmulative squared REsiduals to check the absence of vertical jumps called "outliers"; b) Mean Absolute Deviation; c) Mean Square Error.

References

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis: Theory Application to Travel Dem. MIT Press

MAChuo, K.; Saito, M. 2009. Applicability of FHWA Crash Prediction Module to Selecting Candidate Locations for Safety Audits of Two-Lane Rural Highways, *Transportation Research Record* 2137: 20–28

Cafiso, S., D'Agostino, C., 2016. Assessing the stochastic variability of the Benefit-Cost ratio in roadway safety management. Accident Analysis & Prevention, 93: 189-197. DOI: 10.1016 / j.aap.2016.04.027.

Dell'Acqua, G. 2011. Reducing Traffic Injuries Resulting from Excess Speed Low-Cost Gateway Treatments in Italy, *Transportation Research Record* 2203: 94−99. http://dx.doi.org/10.3141/2203-12.

Dell'Acqua, G. 2015. Modeling Driver Behavior by Using the Speed Environment for Two-Lane Rural Roads, *Transportation Research Record* 2472: 83−90. http://dx.doi.org/10.3141/2472-10.

El-Basyouny, K.; Sayed, T. 2006. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models, *Transportation Research Record* 1950: 9-16. DOI: 10.3141/1950-02

Hough JA, Cao X, Handy SL. Exploring Travel Behavior of Elderly Females in Rural and Small Urban North Dakota An Ecological Modeling Approach. Transp Res Rec. 2008; 2082: 125-131.

Hu, J.-bi; Yang, Y. 2011. Safety of Driving Behavior on Low-Volume Roads in China Case Study of Qinghai–Tibet Highway, *Transportation Research Record* 2203: 100–105. DOI: 10.3141/2203-13.

Islam S, Mannering F. Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence. In J Safety Res. 2006; 37: 267-276.

Kweon, Y.-J.; Oh, C. 2011. Identifying Promising Highway Segments for Safety Improvement Through Speed Management, *Transportation Research Record* 2213: 46–52. DOI: 10.3141/2213-07

McCartt AT, Mayhew DR, Braitman KA, Ferguson SA, Simpson HM. Effects of Age and Experience on Young Driver Crashes: Review of Recent Literature. Traffic Inj Prev. 2009; 10(3): 209-219.

McFadden, D., 1977. An Application of Diagnostic Tests for the Irrelevant Alternatives Property of The Multinomial Logit Model. Transportation Research Record, No. 637, pp. 39–46

Qin, X. 2012. Quantile Effects of Causal Factors on Crash Distributions, *Transportation Research Record* 2279: 40-46. DOI: 10.3141/2279-05

Rosenbloom S. Is the Driving Experience of Older Females Changing? Safety and Mobility Consequences over Time. Transp Res Rec. 2006; 1956: 127-132.

Russo F.; S.A. Biancardo S.A.; Dell'Acqua G. 2014. Road Safety from the Perspective of Driver Gender and Age as Related to the Injury Crash Frequency and Road Scenario. Traffic Injury Prevention. 15(1): 25-33 DOI: 10.1080/15389588.2013.794943.

Russo, F.; Busiello, M.; Dell'Acqua, G. 2016. Safety performance functions for crash severity on undivided rural roads. Accident Analysis and Prevention, 93: 75-91. DOI: 10.1016/j.aap.2016.04.016.

Srinivasan, R.; Baek, J.; Carter, D.; Persaud, B.; Lyon, C.; Eccles, K.; Gross, F.; Lefler, N. 2009. *Safety Evaluation of Improved Curve Delineation*, Report No. FHWA-HRT-09-045, Federal Highway Administration, Washington DC.

Train, K.. Discrete choice methods with simulation. Cambridge university press, 2003.