Accident Analysis and Prevention 42 (2010) 1072–1079

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables

Salvatore Cafiso[a,1], Alessandro Di Graziano[a,2], Giacomo Di Silvestro[a,3], Grazia La Cava[a,4], Bhagwant Persaud[b,*]

[a] Department of Civil and Environmental Engineering, University of Catania, Viale Andrea Doria 6, 95125 Catania, Italy
[b] Department of Civil Engineering, Ryerson University, Toronto, Canada

ARTICLE INFO

Article history:
Received 5 October 2009
Received in revised form 8 December 2009
Accepted 11 December 2009

Keywords:
Road safety
Local rural roads
Accident modeling
Safety performance function
Empirical Bayes

ABSTRACT

In Europe, approximately 60% of road accident fatalities occur on two-lane rural roads. Thus, research to develop and enhance explanatory and predictive models for this road type continues to be of interest in mitigating these accidents. To this end, this paper describes a novel and extensive data collection and modeling effort to define accident models for two-lane road sections based on a unique combination of exposure, geometry, consistency and context variables directly related to the safety performance. The first part of the paper documents how these were identified for the segmentation of highways into homogeneous sections. Next, is a description of the extensive data collection effort that utilized differential cinematic GPS surveys to define the horizontal alignment variables, and road safety inspections (RSIs) to quantify the other road characteristics related to safety. The final part of the paper focuses on the calibration of models for estimating the expected number of accidents on homogeneous sections that can be characterized by constant values of the explanatory variables.

Several candidate models were considered for calibration using the Generalized Linear Modeling (GLM) approach. After considering the statistical significance of the parameters related to exposure, geometry, consistency and context factors, and goodness of fit statistics, 19 models were ranked and three were selected as the recommended models. The first of the three is a base model, with length and traffic as the only predictor variables; since these variables are the only ones likely to be available network-wide, this base model can be used in an empirical Bayesian calculation to conduct network screening for ranking "sites with promise" of safety improvement. The other two models represent the best statistical fits with different combinations of significant variables related to exposure, geometry, consistency and context factors. These multiple variable models can be used, with caution, and in conjunction with results from other studies, to derive accident modification factors for these variables for design applications, and in safety assessment for smaller samples of sites for which these variables can be assembled with relative ease.

1. Introduction

Rural road safety accounts for a considerable share of the total road safety problem. Statistics show that deaths on rural roads account for between 28% and 87% of all road fatalities in different EU countries. In 2006, there were 17,965 fatalities in 18 EU Countries on "non-freeway roads" in rural areas, accounting for 59% of all road fatalities. (European Road Safety Observatory, 2007; European Union Road Federation, 2008).

In Italy, local two-lane rural roads had 1189 fatalities in 2006, accounting for 21% of all road fatalities and 37% of the rural road fatalities (Italian National Institute of Statistics, 2006). These roads provide access to land and towns and are characterized by a low-medium traffic flow (usually in the range 1000–8000 vehicles/day) and short-distance journeys. Considering the complexity of the surrounding environment and the available funds, safety improvements on these roads must be based on a comprehensive approach that takes into account the relative safety performance of different road segments. Accident models, also known as safety performance

* Corresponding author. Tel.: +1 416 979 5000x6464; fax: +1 416 979 5122.
E-mail addresses: dcafiso@dica.unict.it (S. Cafiso), adigraziano@dica.unict.it (A. Di Graziano), gdisilve@dica.unict.it (G. Di Silvestro), glacava@dica.unict.it (G. La Cava), bpersaud@ryerson.ca (B. Persaud).
[1] Tel.: +39 0957382213; fax: +39 0957382247.
[2] Tel.: +39 0957382215; fax: +39 0957382247.
[3] Tel.: +39 0957382211; fax: +39 0957382247.
[4] Tel.: +39 0957382223; fax: +39 0957382247.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.12.015

functions, provide a direct method to analyze how safety performance is related to different road features and characteristics of a road segment.

In the safety assessment of this road class, a considerable problem is the difficulty of getting a large enough sample of accidents, given the low traffic volumes and the correspondingly low accident frequencies, a problem compounded by the fact that in Italy, only injury accidents are recorded in the database. Also, in order to develop and use accident models for these assessments, there is a need to define homogeneous sections with minimum lengths and exposure levels, as noted by Resende and Benekohal (1997) who highlighted the importance of dividing roadways into segments with homogenous characteristics related to both geometry and traffic flow.

A literature review revealed several research efforts to estimate accident models for two-lane rural roads sections (Pardillo and Llamas, 2003; Abdel-Aty and Radwan, 2000; Cafiso et al., 2008b; Garber and Ehrhart, 2000; Zhang and Ivan, 2005). Garber and Ehrhart (2000) considered simply road sections between two major junctions and, using a multivariate ratio of polynomials, found the variables “standard deviation of the speed profile” and “flow per lane” to be strongly correlated to crash rate. Zhang and Ivan (2005) calibrated negative binomial regression models for a sample of segments of constant 1 km length and found a significant relationship between accidents and speed limit, “the sum of the change rate of the horizontal curvature”, “the sum of the change rate of the vertical curvature” and “the maximum horizontal curvature rate”. Pardillo and Llamas (2003) proposed negative binomial regression models for two definitions of segments: (1) 1-km fixed length segments, and (2) network links joining two consecutive nodes with variable lengths ranging from 3 to 25 km. They found significant correlations between accidents and access density, average sight distance, average speed limit and proportion of no passing zones, revealing a need to use these variables in defining homogeneous sections with suitable lengths not less than 400 m. Abdel-Aty and Radwan (2000) divided a 227 km long road into 566 segments with homogeneous characteristics in terms of traffic flow and geometry (degree of horizontal curvature, shoulder and median widths, rural/urban classification, lane width and number of lanes), all variables found to be strongly related to the accident occurrence. The research concluded that to obtain a reliable accident prediction model, sections should be 0.8 km or longer. Perhaps the most significant US research of late has been the calibration of models that formed the basis for those applied in the Interactive Highway Safety Design Model (Paniati, 1996) and, lately, the Highway Safety Manual due for release in 2010. That research (Bared and Vogt, 1998) related accidents per unit of exposure (defined as the product of traffic volume and segment length) to lane width, shoulder width, roadside hazard rating, horizontal curvature, vertical grade and curvature, and driveway density. Another IHSDM related effort (Fitzpatrick et al., 2000) found that the average radius of horizontal curves for a roadway section shows promise as design consistency measure, but does not appear to be as appropriate a measure of design consistency as the speed reduction for individual horizontal curves.

This paper builds on the previous research in developing accident models for network screening analysis and for evaluating the influence of some key road features on safety. The research approach is novel in that it involved the selection of, and assembly of data on, explanatory variables related to exposure, geometric design, design consistency and roadside features, a combination that has not been addressed previously. These variables include a number of key explanatory factors not typically available in routinely available databases or not previously explored. The segmentation of a 168.20 km sample of roads used for the model calibration is a relatively new approach that is based on a methodology detailed by Cafiso et al. (2008b) for dividing a two-lane road sample into segments characterized by homogeneous highway features related to safety.

2. Data survey

2.1. Road data

The survey was conducted on a sample of 168.20 km of two-lane local rural roads located in Italy. Based on a literature review (Pardillo and Llamas, 2003; Zhang and Ivan, 2005; Cafiso et al., 2006, 2008a; Bared and Vogt, 1998), and on the authors' experience (Cafiso et al., 2008b), the features identified as the base parameters for a safety performance evaluation were curvature (radius, length), tangent length, cross section (lane and shoulder width and type), density of driveways, and roadside hazard.

Other information were considered (e.g., sight distance, gradient, pavement condition) but the above-listed parameters are particularly relevant and were relatively easy to collect. Due to paucity of as-built drawings of local rural roads in Italy, an alternative, cost effective and practical method for collecting road data was required.

A GPS survey was used to collect horizontal alignment information (curvature and tangent length) and road safety inspections (RSIs) (Cafiso et al., 2008a,b, 2007a) were conducted in order to quantify the characteristics of the other feature (cross section, density of driveways, and roadside hazard). The GPS investigation was performed by driving a car along the centre-line of the lane at a moderate speed (60 km/h), using L2 GPS equipment in cinematic-differential mode (DGPS) so as to obtain an error of less than 10 cm. This surveying procedure allows a series of points to be identified along the trajectory of the vehicle (lane axis) for a base representation of the stretch, but it is not capable of recognizing the single parameters of the alignment design. Therefore, the next phase of the procedure aimed to identify the single elements of the lane axis (tangents, circular curves, clothoids) and determine their values (length, radius, angular extension). This procedure uses regression splines with the aid of suitable smoothing factors to correct measurement errors or those resulting from an actual vehicle trajectory that strays from the lane axis. More details of the survey method and spline smoothing procedure are in Cafiso et al. (2008b) and Cafiso and Di Graziano (2008).

Roadside hazard and driveway density were measured using a new operating inspection procedure that is capable of improving the effectiveness and reliability of the RSIs application (Cafiso et al., 2006, 2007a,b, 2008a).

2.2. Accident data

A 5-year analysis period was chosen for the investigation period to compensate for the low traffic flow and accident frequencies usually expected on local rural roads. Data for accidents with at least one fatality or injury were acquired from police reports. These data included location reference information and kilometre markers. Accidents located in built-up areas and at intersections were not used. Data for 279 accidents were collected for the 5-year period chosen, with a total of 640 injured persons and 16 fatalities.

3. Data treatment

On the basis of the data collected, the following parameters were identified in order to divide the sample into homogeneous sections:

- Average Annual Daily Traffic (AADT) to describe the exposure to accident risk.

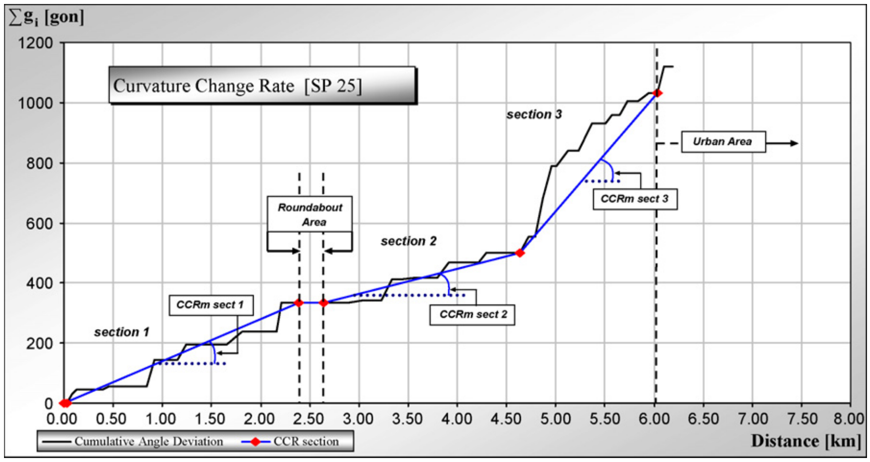

Fig. 1. Evaluation of homogeneous sections based on the cumulative curvature change rate of the section CCR_{sect}.

- Curvature Change Rate (CCR) and average paved width (W) to describe main road geometry characteristics.
- Road Side Hazard rating (RSH) to describe roadside conditions.

While it is obvious that AADT and W can be directly measured, both CCR and RSH needed to be determined from parameters characterizing the segment. To identify the sections with constant CCR a *Section Curvature Change Rate* (CCR_{sect}) was defined as follows:

$$CCR_{sect} = \frac{\Sigma_i|\gamma_i|}{L}\ [\text{gon/km}] \tag{1}$$

where γ_i is the deflection angle for a contiguous element (curve or tangent) i within a section of length L. Along an entire road corridor, the sum of the deflection angles γ_i [gon: centesimal degree] can be represented in a diagram as a function of distance from the start of the corridor (Fig. 1). Road sections with homogenous horizontal alignment are identified in this graph by "trend" lines, where the slope of the cumulative angle deviation curve is relatively constant, as shown in Fig. 1. Based on the German procedure (RAS-L, 1995) a minimum length of the sections of 2 km can be adopted. The CCR value (CCR_{sect}), for an identified section is equal to the slope of a smoothed line representing the section, as shown in Fig. 1.

The *Roadside Hazard* (RSH) was evaluated using specially designed checklists filled in by inspectors for 200 m segments. Road Safety Inspections were carried out for both directions (right and left side of the road) assigning a score (0 = not present, 1 = low risk, 2 = high risk) to 5 different roadside items based on previous studies (Cafiso et al., 2006, 2007a,b, 2008a). The relative weights for specific roadside safety hazard items were 3 for embankments, 5 for bridges, 2 for dangerous terminals and transitions, 2 for trees, utility poles and rigid obstacles, and 1 for ditches. To define an overall value of RSH_i for each segment, a weighted mean of the five roadside items was calculated as follows:

$$RSH_i = \frac{\sum_{k=1}^{2} \max(Score_{ijk} \times Weight_j)}{2} \tag{2}$$

where k is the direction of the survey (1 = right side; 2 = left side); $Score_{ijk}$ is the score (0, 1, or 2) of the roadside safety items j (j = 1, 2, ..., 5) in the ith inspection unit along the direction k; $Weight_j$ = relative weight of the jth roadside safety item based on the AASHTO severity indices (AASHTO, 1996).

With Eq. (2), the values {0, 1, 2} assigned during the Road Safety Inspection for each of the 5 different roadside hazard items were used to define a RSH score on a scale from 0 to 10.

The RSH rating describes roadside conditions based on the judgment of an experienced technical team, so there is some subjectivity in evaluating this variable. This Road Safety Inspection procedure was tested in a previous study (Cafiso et al., 2006), but even if the subjectivity in the evaluation of RSH was reduced it remains implicit in its definition.

Homogeneous road sections, in which a sequence of RSH values can be considered constant, were identified by minimizing the sum of squared deviations of the individual RSH values with respect to the mean and conducting a t-test in this process. In order to avoid very short homogeneous sections with respect to the RSH segmentation, a minimum length of 1000 m and a t-test significance level of 15% were selected. A lower level of significance for the t-test in RSH segmentation is not recommended in order to avoid the selection of contiguous short sections with similar RSH average values (Cafiso et al., 2008b).

The final definition of a homogeneous road section is a section where all the above-mentioned parameters (CCR_{sect}, W, AADT, mean RSH) are constant. In Fig. 2, six homogeneous sections from an example 10-km road taken from the investigated sample are shown. In total, 107 homogeneous sections were defined on a total of 168.20 km of roads. The length of these sections ranges from 0.50 to 4.29 km.

4. Selection of explanatory variables

Once the homogeneous sections were defined, additional variables could be assigned to the road sections in order to estimate a model. These variables could be relevant in explaining the accident occurrence and distribution but their variability does not influence the homogeneity of the section because they are derived from the basic descriptive parameters used to define the segmentation. These additional variables can be related to the geometric and operational conditions and to the design consistency of the road as follows.

With regard to geometric and operational variables (Pardillo and Llamas, 2003) the *Curvature Ratio* (CR) and *Tangent Ratio* (TR) were computed using the following equations:

$$CR = \frac{\sum_{j=1}^{k} L_{Cj}}{L_{HS}} \tag{3}$$

$$TR = \frac{MAX_{e=1}^{t}(L_{Te})}{L_{HS}} \tag{4}$$

where L_{HS} is the total length of the homogeneous section [km]; L_{Cj} is the length of jth curve in the homogeneous section composed by k curves [km]; L_{Te} is the length of eth tangent in the homogeneous section composed by t tangent [km].

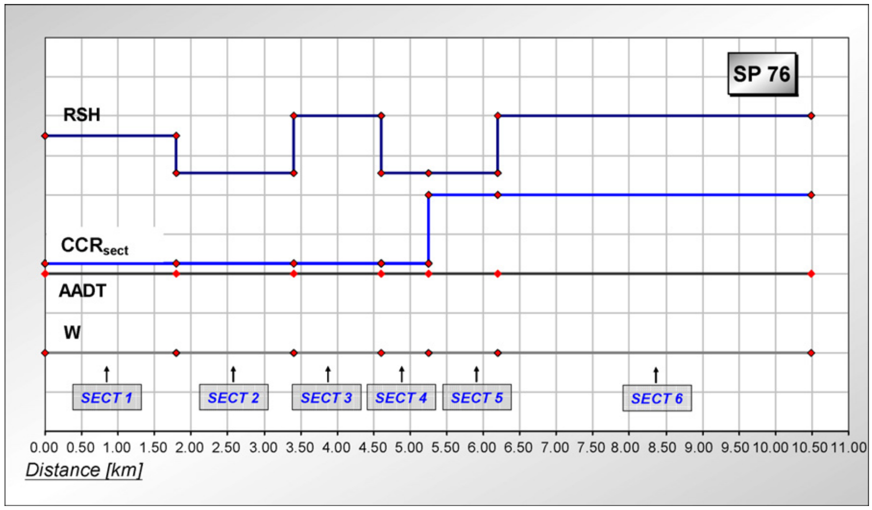

Fig. 2. Illustration of homogeneous section segmentation.

The *Average Operating Speed* (V_{avg}) was computed on the basis of the operating speed profile as follows:

$$V_{avg} = \frac{\sum_{i=1}^{n} V_{85i} \cdot L_i}{L_{HS}} \quad (5)$$

where V_{85i} is the operating speed of ith geometric element [km/h], L_i the ith element length of the homogeneous section [km] and n is the number of geometric elements along a section.

The operating speeds (V_{85}) on tangents and curves were calculated using a prediction model specifically developed for two-lane local rural roads in Italy (Cafiso et al., 2007b):

$$V_{85i} = \frac{0.987 \times V_{env} - 0.0418\mathrm{CCR_S}i \times V_{env}}{100} \;[\mathrm{km/h}] \quad (6)$$

where the environmental speed of road section (V_{env}) is calculated as follows:

$$V_{env} = 100.05 - 0.197\mathrm{CCR_{sect}} + 2.147W \quad (7)$$

and the Curvature Change Rate of ith element is calculated as before as follows:

$$\mathrm{CCR_S}i = \frac{g_i}{L_i} \;[\mathrm{gon/km}] \quad (8)$$

where g_i is the deflection angle of the ith geometric element [gon].

Then, in order to define a reliable operating speed profile, it was assumed, based on Poulos and Mattar-Habib (2004), that the acceleration and deceleration phases between tangent and curve lasted for 2 s on the tangent and for 1 s on the curve.

As design consistency variables, the two measurements, relative area (R_a) bounded by the speed profile and the average operating speed (V_{avg}), and standard deviation (σ) of the operating speed profile, were calculated from the following equations (Poulos and Mattar-Habib, 2004):

$$R_a = \frac{\sum_{i=1}^{n} a_i}{L_{HS}} \;[\mathrm{m/s}] \quad (9)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (V_{85} - V_{avg})^2}{n}} [\mathrm{km/h}] \quad (10)$$

where a_i is the area bounded by the operating speed profile and the average operating speed line [m^2/s]; V_{85i} is the operating speed of the ith geometric element [km/h]; V_{avg} is the average operating speed along the entire homogeneous section of length L_{HS} [km/h]; n is the number of geometric elements in the homogeneous section.

Three more consistency variables were selected considering the speed differentials ($\Delta V_s = |V_{85i} - V_{85i+1}|$) between contiguous elements in the homogeneous section, using the following equations:

$$\Delta V_n = \frac{\sum_{s=1}^{n_{\Delta V}} \Delta V_s}{n_{\Delta V}} \;[\mathrm{km/h}] \quad \text{average speed differential} \quad (11)$$

$$\Delta V_{10} = \frac{N(\Delta V > 10)}{L_{HS}} [1/\mathrm{km}] \; \Delta V_{10} \text{ density} \quad (12)$$

$$\Delta V_{20} = \frac{N(\Delta V > 20)}{L_{HS}} \;[1/\mathrm{km}] \; \Delta V_{20} \text{ density} \quad (13)$$

where $n_{\Delta V}$ is the number of speed differentials (ΔV_s) in the homogeneous section (HS); $N(\Delta V > 10)$ is the number of speed differentials(ΔV_s) higher than 10 km/h in the HS; $N(\Delta V > 20)$ is the number of speed differentials (ΔV_s) higher than 20 km/h in the HS.

Fig. 3 shows an example of how the operating speed profile measurements are related to horizontal alignment.

With respect to the context-related variables, since it has been demonstrated that direct accesses to roads can significantly increase accidents (Miaou and Lum, 1993), driveway density (DD) was considered and obtained from the RSI checklists previously described, using the following equation:

$$\mathrm{DD} = \frac{\text{number of accesses in the HS (both directions)}}{L_{HS}} \quad (14)$$

5. Accident modeling

5.1. Regression technique and model form

The models proposed are based on the Generalized Linear Modeling approach (GLM), which has the advantage of overcoming the limitations of conventional linear regression in accident frequency modeling. In particular, it facilitates the assumption of a Negative Binomial error structure, which is more pertinent to accident frequency variation.

The general form of the accident prediction model adopted is:

$$E(Y) = e^{a_0} \cdot L \cdot AADT^{a_1} \cdot e^{\sum_{j=1}^{m} b_j x_j} \quad (15)$$

where $E(Y)$ is the expected injury accident frequency/year; L is the length of the segment under consideration (km); AADT is the Average Annual Daily Traffic (AADT) (veh/day); x_j is the any of m-additional variables; a_0, a_1, and b_j are the model parameters.

This model form was selected because it is generally accepted (Pardillo and Llamas, 2003; Abdel-Aty and Radwan, 2000; Cafiso et

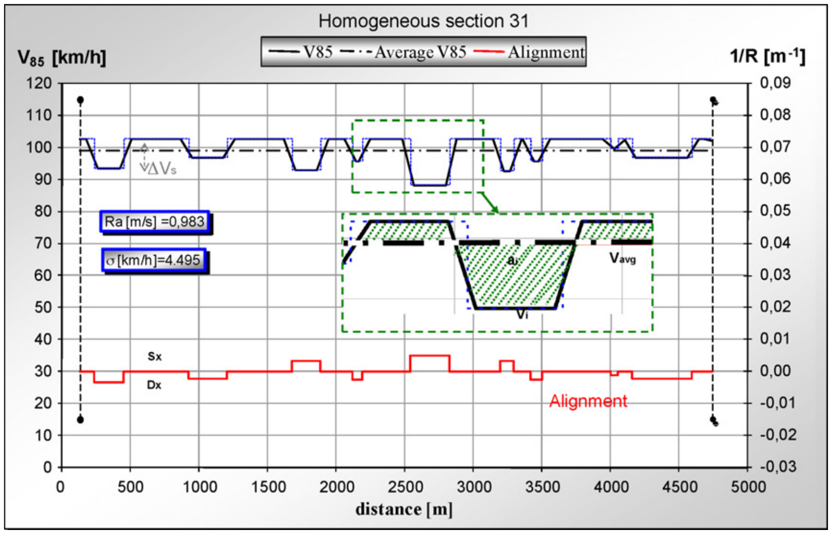

Fig. 3. Operating speed profile and related variables.

al., 2008b; Zhang and Ivan, 2005). In particular, it logically estimates zero accidents if one of the two exposure variables (AADT or L) is equal to zero.

To evaluate the model form, cumulative residual analysis (Hauer, 2004; Lord and Persaud, 2000) was performed. The residual for each section of the sample is equal to the difference between the observed and estimated values of the dependent variable. A standardized residual for the ith section (SR_i) was computed from the following equation:

$$\mathrm{SR}_i = \frac{(y_i - \hat{y}_i)}{\sqrt{\hat{y}_i + k \cdot \hat{y}_i^2}} \tag{16}$$

where y_i is the observed number of accidents of the ith segment; $\hat{y}_i$ is the estimated number of accidents of the ith segment; k is the overdispersion parameter.

The cumulative standardized residual value for the jth element is obtained as follow:

$$\mathrm{CSR}_j = \sum_{i=1}^{j} \mathrm{SR}_i \tag{17}$$

for $j = 1, \ldots, m$ and with m equal to the total number of the observation.

Cumulative standardized residual analysis was carried out by plotting CSR_i versus exposure values, computed as the product of AADT and the length for each homogeneous section. The closer the CSR curve stays to the x-axis, avoiding significant increases or decreases in ordinate, the more appropriate is the model form chosen (Hauer and Bamfo, 1997). A CSR curve that is contained within 2 standard deviations is considered satisfactory.

Three measurements were used to assess the goodness-of-fit of the model following Sutterhwaite (1981) and McCullagh and Nelder (1989). These were the Pearson χ^2 statistics the Akaike's Information Criterion (AIC) and an adjusted R^2 value.

The Pearson χ^2 can be calculated by means of the following formula:

$$\text{Pearson } \chi^2 = \sum_{i=1}^{n} \frac{\left[y_i - \hat{E}(y_i)\right]^2}{\mathrm{Var}(y_i)} \tag{18}$$

where $\mathrm{Var}(y_i)$ is the variance of the ith segment.

For a model to be acceptable, the t-ratio for the model parameters and the Pearson χ^2 of the model must be less than a critical value of χ^2 distribution value that is based on the model's degrees of freedom and a level of significance of α.

The Akaike's Information Criterion (AIC) has been extensively used as goodness of fit measure (Abdel-Aty and Radwan, 2000; Zhang and Ivan, 2005; Montella et al., 2008). In contrast to the Pearson χ^2 test, the AIC is not strictly a test of the model acceptability; rather it is a means for comparing models, i.e., a tool for model selection.

The AIC value is calculated as follows:

$$\mathrm{AIC} = -2\log L + 2p \tag{19}$$

where $\log L$ is the maximum log-likelihood of the fitted model; p is the number of parameters in the model.

The smaller the value of AIC, the better is the model data fit. Given a dataset, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

The adjusted R^2 value (Miaou, 1996) yields a value between 0 and 1, and is explicitly based on the overdispersion parameter and the number of explanatory terms in a model and so is especially applicable in evaluating negative binomial models. Its value is computed as follows:

$$R_k^2 adj = 1 - \left(\frac{k_0}{k}\right) \cdot \frac{(n-1)}{(n-p-1)} \tag{20}$$

where n is the total number of observations in the sample (107 for this study); p is the total number of the parameters in the model (not counting the constant term); k_0 is the overdispersion parameter estimated in the negative binomial model with only a constant term; k is the negative binomial overdispersion parameter estimated in the full model. $R_k^2 adj$ is also useful to compare the goodness-of-fit of models calibrated on different datasets.

5.2. Model development

Table 1 shows summary statistics of the variables derived from the data and considered for model development. The final selection of variables was based on an initial analysis to identify the correlation between two independent variables among those reported in Table 1. The most common measure of correlation is the Pearson Product Moment Correlation (Pearson's correlation (CP)) (Pearson, 1896) which reflects the degree of linear relationship between two variables, with values of +1 or −1 indicating perfect positive or negative linear correlation, respectively.

Table 1
Summary of characteristics of homogeneous section variables considered for model development.

Variable group	Abbreviation	Description	Mean	Min.	Max.	Standard deviation
Exposure	L_{HS} [km]	Length of homogeneous section	1.57	0.50	4.29	0.83
	AADT [veh/day]	Average Annual Daily Traffic	3811	600	12,400	2520
Geometric and operational	CCR_{sect} [gon/km]	Section curvature change rate	114.79	0.00	629.21	128.85
	W [m]	Paved width	7.33	5.70	10.00	1.22
	TR	Tangent ratio	0.38	0.04	1.00	0.27
	CR	Curve ratio	0.31	0.00	0.71	0.18
	V_{avg} [km/h]	Average operating speed	92.8	57.5	115.5	15.5
Consistency	R_a [m/s]	Relative area bounded by the speed profile	1.23	0.00	4.06	0.95
	σ [km/h]	Standard deviation of operating speed profile	6.98	0.00	18.46	5.13
	ΔV_n [km/h]	Average speed differential	10.16	.00	28.63	6.68
	ΔV_{10} [no./km]	Speed differentials density (higher than 10 km/h) in the Homogeneous section	2.94	0.00	14.30	3.53
	ΔV_{20} [no./km]	Speed differentials density (higher than 20 km/h) in the homogeneous section	1.40	0.00	12.55	2.30
Context-related	RSH	Roadside hazard	1.52	0.00	4.93	1.07
	DD[n driveways/km]	Driveway density	6.77	0.80	21.81	4.64

Variable pairs that are strongly correlated (CP higher than ±0.85 and P-value below 0.05 indicating statistically significant non-zero correlations at the 95% confidence level) were discarded. Specifically, there were significant correlations between ΔV_{10} and ΔV_{20}, ΔV_{10} and R_a, σ and ΔV_n, σ and R_a, V_{avg} and R_a, and between V_{avg} and CCR_{sect}.

Based on the results of the Pearson's Correlation analysis, and on logical considerations, the following 8 sets of non-correlated variables were identified for model development:

(1) CR–DD–RSH–TR–W–CCR_{sect}–ΔV_{10}–ΔV_n
(2) CR–DD–RSH–TR–W–CCR_{sect}–ΔV_{10}–s
(3) CR–DD–RSH–TR–W–CCR_{sect}–ΔV_{20}–s
(4) CR–DD–RSH–TR–W–ΔV_{10}–ΔV_n–V_{avg}
(5) CR–DD–RSH–TR–W–ΔV_{10}–s–V_{avg}
(6) CR–DD–RSH–TR–W–ΔV_{20}–ΔV_n–V_{avg}
(7) CR–DD–RSH–TR–W–ΔV_{20}–s–V_{avg}
(8) CR–DD–RSH–TR–W–CCR_{sect}–ΔV_{20}–ΔV_n–R_a

Models were calibrated for unique combinations of non-correlated variables from the above sets. The next step was to dismiss any model characterized by a Pearson χ^2 goodness-of-fit parameter higher than the critical χ^2 value or, at least, by one variable having a P-value greater than or equal to 0.05. For the roadside hazard (RSH) variable only, a statistical significance level of 10% was considered a reasonable threshold due to the lack of a totally objective way of evaluating the RSH.

5.3. Results

From the 19 models ultimately calibrated, models 1, 15 and 19 were selected as recommended models for reasons outlined below. Table 2 presents estimates of parameters, the corresponding t-statistic, and the goodness of fit measures for these three models.

As seen, all three models have an AADT exponent less than 1, which implies, in accord with other research for two-lane roads, that the accident-AADT relationship is non-linear, with a slope that decreases with increasing AADT. Model 1, a traffic-exposure only based model, can be applied without collecting further information for a road segment. Although this model has a Pearson χ^2 value considerably smaller than the corresponding χ^2 value, it does have a high AIC value, likely because of the absence of non-exposure variables. However Model 1 can be applied in an Empirical Bayesian procedure to evaluate potential for safety improvement (FHWA, 2002) in conducting network screening analysis since non-exposure variables are unlikely to be readily available for the entire network.

Models 15 and 19 have the best statistical fit. The Pearson χ^2 values are smaller than the corresponding critical χ^2 critical values, and they also have the lowest AIC values of all 19 models (both equal to 407.40).

Fig. 4 shows the plot of the cumulative standardized residuals (CSR) curve just for Model 19, the model that best fits the observed data. This plot reveals the appropriateness of the model form for the representation of the exposure variable. CSR plots for Models 1 and 15 show similar results.

Table 2
Model parameters and goodness-of-fit indicators MODEL FORM : $E(Y) = e^{a_0} \cdot L_{HS} \cdot AADT^{a_1} \cdot e^{\sum_{j=0}^{m} b_j x_j} \left[\frac{\text{acc}}{\text{year}}\right]$.

Model	Parameter estimate and *t-statistic*						k	AIC	Pearson χ^2	$\chi^2_{0.05}$	$R^2_k adj.$
	a_0	a_1	b_0	b_1	b_2	b_3					
1	CONST −7.132 *<.001*	AADT 0.731 *<.001*					0.93	417.6	85.4	129.9	0.31
15	CONST −6.682 *<.001*	AADT 0.619 *<.001*	DD 0.0646 *0.006*	CR −1.89 *0.011*	s 0.0691 *0.004*		0.66	407.4	106.1	126.6	0.50
19	CONST −7.812 *<.001*	AADT 0.753 *<.001*	DD 0.067 *0.004*	CR −1.948 *0.009*	ΔV_{10} 0.0872 *0.023*	RSH 0.185 *0.07*	0.71	407.4	97.9	125.5	0.45

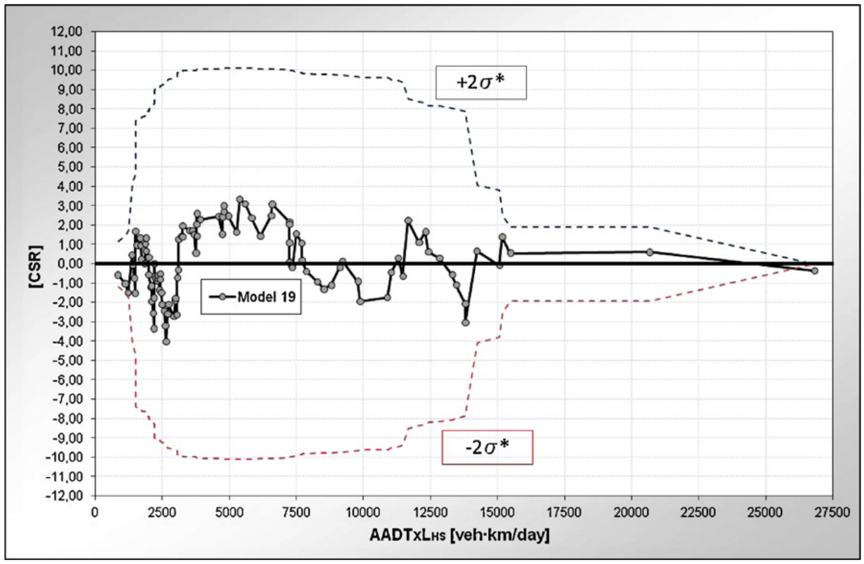

Fig. 4. Cumulative standardized residuals plot for exposure for Model 19.

Models 15 and 19 have practical engineering value as safety performance indicators because at least one significant parameter related to exposure, geometric, consistency and context factors, is included in the models. They can be used to evaluate the safety performance of design alternatives related to the features highlighted in the models and/or to identify specific safety-related problems in order to optimize and target safety improvements at existing sites.

Model 15 includes the exposure variables segment length and AADT, but also captures the influence of the context-related variable, driveway density, the consistency variable, σ, representing the standard deviation of speed, and the geometric variable, curvature ratio. It is worthy of note that the parameter estimates of σ and curvature ratio are, respectively, positive and negative. This means that with increasing standard deviation of speed (σ) an increase in accidents can be expected and with increasing curvature ratio (CR) a decrease in accidents can be expected. These results can be regarded as intuitively sound considering, for example, that increased speed variation can result from unsafe design situations such as the transition from a long tangent to a sharp curve; and drivers may tend to be more vigilant and drive more carefully on segments with larger curvature ratios.

Model 19 also captures the influence of the context-related variables, driveway density (DD) and roadside hazard (RSH), the consistency variable, DV_{10}, representing the density of speed differentials higher than 10 km/h in a homogeneous section, and the geometric variable, curvature ratio (CR). As was the case for model 15, the parameter estimate of curvature ratio is negative; moreover it can be noted from the ΔV_{10} parameter estimate that an increase in the number of speed differentials higher than 10 km/h is associated with an increase in the expected number of accidents. These results, too, can be considered to be intuitively reasonable from an engineering point of view. Finally, for both models 15 and 19, an increase in the value of the context-related variables (driveway density for model 15, driveway density and roadside hazard for model 19) is associated with an increase in accident frequency.

6. Conclusions

Due to the typically low traffic volume and to the lack of a significant number of accidents on two-lane local rural roads, it is necessary to divide these roads into homogeneous sections with a minimum length in order to develop reasonable accident models. Thus, for each road in the sample assembled in this study to develop such models, variables that can be related to safety along a homogeneous section were identified and collected. The result is that a comprehensive procedure including low-cost data surveying, road segmentation methodology and model development is presented, specifically for local rural roads.

14 variables belonging to four main groups (exposure, geometric, consistency and context) were identified and used to estimate several models using the Generalized Linear Modeling approach with a Negative Binomial error structure. Three models were selected as recommended models based on practical considerations, statistical significance, and on goodness of fit indicators:

- Model 1 includes only the exposure variables, length (L_{HS}) and traffic volume (AADT).
- Model 15 includes length (L_{HS}), traffic volume (AADT), driveway density (DD), curvature ratio (CR) and the standard deviation of the operating speed profile (s).
- Model 19 includes length (L_{HS}), traffic volume (AADT), driveway density (DD), roadside hazard rating (RSH), curvature ratio (CR) and number of speed differentials higher than 10 km/h (ΔV_{10}).

Model 1 can be applied in an Empirical Bayes procedure to conduct network screening analysis since non-exposure variables are unlikely to be readily available for the entire network. Models 15 and 19, which have the best fits and at least one variable pertaining to the four main groups of variables (exposure, geometric, consistency and context), can be used to evaluate the safety performance of existing roads or alternative design improvement solutions.

From a pragmatic perspective, Model 19 may be preferable to Model 15 since, in addition to the driveway density variable, it includes roadside hazard (a variable evidently related to the accident severity based on previous research). It is also easier to apply than Model 15, which requires the determination of the standard deviation of speed (s) based on an operating speed profile. By contrast, for the computation of the number of speed differentials higher than 10 km/h (ΔV_{10}) for Model 19, it is enough to obtain the measurements between contiguous elements (curves and tangents) along a homogeneous section.

The models are all relatively lean, but can still be used for empirical Bayes (EB) accident prediction applications, since unexplained variation due to omitted variables is explicitly considered in the

EB procedure. Numerical information on accident modification factors (AMFs) may also be derived from the model coefficients, but caution in doing so is recommended since this is in effect a cross sectional analysis and, as Hauer (2004) suggests, implied AMFs need to be assessed for consistency with those derived in other studies using independent data. This type of assessment has of late been performed by specially convened expert panels (Washington et al., 2009) and in systematic reviews (Elvik, 2005; Bahar, 2009).

Acknowledgements

This research is part of a project (PRIN 2006) co-financed by the Italian Ministry of University and Research whose support is gratefully acknowledged. The authors would like to also acknowledge the Road Agencies of the Provinces of Catania, Siracusa and Ragusa for supporting the data acquisition. Collaboration between Canadian and Italian researchers was supported by the Natural Sciences and Engineering Research Council of Canada.

References

AASHTO, 1996. Roadside Design Guide. Washington, DC.
Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis & Prevention 32, 633–642.
Bahar, G., 2009. Methodology for the development and inclusion of accident modification factors (AMFs) in the first edition of the highway safety manual. Draft Transportation Research Circular (January).
Bared, J., Vogt, A., 1998. Accident models for two-lane rural segments and intersections. Transportation Research Record: Journal of the Transportation Research Board 1635, 18–29.
Cafiso, S., Di Graziano, A., 2008. Automated in-vehicle data collection and treatment for existing roadway alignment. In: 4th International Gulf Conference on Roads Compendium of Papers, CD-ROM. Doha, Qatar.
Cafiso, S., La Cava, G., Montella, A., 2007a. Safety index for evaluation of two-lane rural highways. Transportation Research Record 2019, 136–145.
Cafiso, S., La Cava, G., Montella, A., Di Graziano, A., Pappalardo, G., Taormina, S., 2007b. Final research report. Identification of hazard location and ranking of measures to improve safety on local rural roads: EU IASP project http://ec.europa.eu/transport/roadsafety/publications/projectfiles/iasp_en.htm.
Cafiso, S., La Cava, G., Montella, A., Pappalardo, G., 2008a. Operative procedures for safety inspections on two-lane rural roads. In: ARACNE, Rome, Italy.
Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., 2008b. Safety performance indicators for local rural roads: a comprehensive procedure from low-cost data survey to accident prediction model. In: TRB 87th Annual Meeting Compendium of Papers, CD-ROM, TRB. National Research Council, Washington, DC.
Cafiso, S., La Cava, G., Montella, A., Pappalardo, G., 2006. A procedure to improve safety inspections effectiveness and reliability on rural two-lane highways. The Baltic Journal of Road and Bridge Engineering 1 (3), 143–150.
Elvik, R., 2005. Introductory guide to systematic reviews and meta-analysis. Transportation research record. Journal of the Transportation Research Board 1908, 230–235.
European Road Safety Observatory, 2007. Annual statistical report 2006. Safety Net Deliverable no. D.1.9. Brussels.
European Road Statistics, 2008. European Union Road Federation (ERF). http://www.irfnet.eu/en/2008-road-statistics/.
Fitzpatrick, K., Elefteriadou, L., Harwood, D., Collins, J., McFadden, J., Anderson, I., Krammes, R., Irizarry, N., Parma, K., Bauer, K., Passetti, K., 2000, August. Speed prediction for two-lane rural highways. Federal Highway Administration. Publication No. 99-171.
FHWA, 2002. SafetyAnalyst: Software Tools for safety management of specific highway sites, white paper for module 1—network screening. http://www.safetyanalyst.org/whitepapers/module1.pdf.
Garber, N.G., Ehrhart, A.A., 2000. Effect of speed, flow and geometric characteristics on crash frequency for two lane rural roads. Transportation Research Record 1717, 76–83.
Hauer, E., 2004. Statistical road safety modeling. Transportation Research Record: Journal of the Transportation Research Board 1897, 81–87.
Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables. In: Proceedings of ICTCT 97 Conference, Lund, Sweden.
Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research Record 1717, 102–108.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, New York.
Miaou, S., Lum, H., 1993. Modelling vehicle accident and highway geometric design relationship. Accident Analysis & Prevention 25, 689–709.
Miaou, S.P., 1996. Measuring the goodness of fit of accident prediction models. Publication FHWA-RD-96-040. FHWA.
Montella, A., Colantuoni, L., Lamberti, R., 2008. Crash prediction models for rural motorways. Transportation Research Record 2083, 180–189.
Paniati, J., 1996. Interactive highway safety design model (IHSDM): designing highways with safety in mind. Transportation Research Circular 453. Transportation Research Board.
Pardillo, J.M., Llamas, R., 2003. Relevant variables for crash rate prediction in Spain's two lane rural roads. In: TRB 82nd Annual Meeting Compendium of Papers, CD-ROM, TRB. National Research Council, Washington, DC.
Pearson, K., 1896. Mathematical contributions to the theory of evolution III. Regression, heredity and Panmixia. Philosophical Transactions of the Royal Society London Series A 187, 253–318.
Poulos, A., Mattar-Habib, C., 2004. New consistency model for rural highways and its relationship to safety. Journal of Transportation Engineering, ASCE 130 (3), 286–293.
RAS-L., 1995. Forschungsgesellschaft für Strassen und Verkehrswesen Linienführung. Richtlinien für die Anlage von Strassen, Bonn, Germany.
Resende, P., Benekohal, R., 1997. Effect of roadway section length on accident modelling traffic congestion and traffic safety. In: The 21st Century Conference, ASCE, Chicago, IL.
Sutterhwaite, S., 1981. A survey of research into relationship between traffic accidents and traffic volumes. In: Supplementary Report 692. Transport Research Laboratory, Crowthorne Berks, UK.
Washington, S., Lord, D., Persaud, B., 2009. Use of expert panels in highway safety: critique. Transportation Research Record: Journal of the Transportation Research Board 2102, 101–107.
Zhang, C., Ivan, J.N., 2005. Effects of geometric characteristics on head-on crash incidence on two-lane roads in Connecticut. Transportation Research Record 1908, 159–164.