

RISK ANALYSIS IN ROAD SAFETY: AN INDIVIDUAL RISK MODEL FOR DRIVERS AND PEDESTRIANS TO SUPPORT DECISION PLANNING PROCESSES

F.A. MARCIANÒ & A. VITETTA

Department of Computer Science, Mathematics, Electronics and Transportation,
Mediterranea University of Reggio Calabria, Italy.

ABSTRACT

Road accident statistics highlight the importance of the number of fatalities and injuries. For example, in Italy, 215,405 road accidents occurred in 2009, resulting in 4,237 deaths and 307,258 injured. One aspect of road safety concerns the protection of ‘vulnerable road categories’ like pedestrians and cyclists. The probability of pedestrians being injured if they are involved in a road accident is higher since, unlike motorized users and cyclists, they do not wear any protection. Current hazard levels could be reduced through interventions targeting recognized network critical points (black spots). Planning must be preceded by risk analysis to support the decisional process through quantitative evaluation. For this reason, we present an individual risk model in the case of a road accident. The purpose is to estimate the probability of a pedestrian being involved in a road accident and the probability of being injured under pre-assigned conditions. The model calculates the individual risk of drivers and pedestrians moving in predefined accident scenarios that can be changed according to the analytical purpose. It also allows quantitative evaluation of how each attribute shares in the accident and how the risk level changes with changing attributes. Attributes are quantitative values that define the accident scenario and pedestrian and driver characteristics. In practice, the model could support the decision planning process, allowing both comparison of hazard levels of various hypothetical scenarios and analysis of the weight and elasticity of each attribute characterizing the accident scenario.

Keywords: drivers, pedestrians, risk analysis, road accident.

1 INTRODUCTION

Risk analysis requires preliminary identification of a risk scenario: in other words, the scenario in which a system structure can be modified by an event. In the transportation field, increasing transport demand has resulted in the growth and development of road networks and an exponential increase of the number of circulating vehicles. Even if this has led to improvements in travel standards and trip comfort, it has also resulted in considerable effects on the environment and urban area liveability. Negative consequences include those of an ordinary nature, such as congestion, delays, and stress for users and road accidents that can be attributed to unusual or unexpected transport system behavior [1].

Risk analysis is the organized and hierarchical analysis of identified causes (class, type, frequency, probability) and effects (type, impact area, measure) of incidental events that might occur during system operation. The risk models presented in the scientific literature, as extensively reported in the next section, have some limitations since they consider probability and vulnerability in an aggregate way. Such limitations do not allow us to evaluate the effect of each attribute and the relative elasticity in the risk value. This knowledge is fundamental, especially in planning operations.

In this paper we present an experimental model to quantify risk level for drivers and pedestrians in the case of a road accident. The purpose is to calibrate risk functions that are applicable in different contexts by entering some external data characterizing the accident scenario. The model allows individual risk value to be estimated for drivers and pedestrians

through a disaggregate probability and vulnerability analysis. The model also estimates the weight and elasticity of each attribute that characterized an accident scenario and it is estimated only for the Italian case. The model is proposed as an effective tool designed for scientific research or policies involved in reducing the number of road accidents. The importance of this aim emerges from the mere fact that in Italy, every day, about 12 persons die and 842 remain injured due to their involvement in a road accident [2].

The proposed model could support the decision planning process, allowing both comparison of hazard levels of various hypothetical scenarios and analysis of the weight and elasticity of each attribute characterizing the accident scenario. Policymakers and designers or researchers can analyze situations and users' behavior at the road accident and then identify the attributes which are necessary to intervene by verifying how risk level varies with changes in the attributes (what-if approach). It is then possible to determine how far each attribute shares in accident occurrence and how risk level varies as the attribute changes.

Thanks to its state-of-the-art approach and analytical structure, the proposed model allows a more detailed analysis since it can estimate the individual risk both of drivers and pedestrians moving in predefined accident scenarios, representing the 'snapshot' of hypothetical real traffic situations, which are obtained both by contextualizing the environment of the road accident and characterizing the users involved.

In Section 2, a literature review and the main objectives of the paper are presented; Section 3 deals with the methodology for specification, calibration, and validation of the proposed model and it is calibrated and validated in Section 4; Section 5 describes a test performed on the city of Reggio Calabria (Italy) to evaluate model transferability and reliability using quantitative statistics. Conclusions and further developments are reported in Section 6.

2 LITERATURE REVIEW

Absolute safety is an abstract concept because it describes a situation with the absence of any damage. In this case, safety is not quantifiable. For this reason, to assess the safety of a system, one should assess 'how far' it is from absolute safety. A measure of this distance represents the risk [1]. Thus risk can be defined as the undesired consequence of an event in relation to its probability of occurring [3] or as the distance between the real context and that of absolute safety [1]. It is a measure of the consequences in terms of deaths and injuries after a road accident has occurred. Such definitions indicate that risk (R) can be calculated through eqn (1), if the event occurrence is estimated through a probabilistic model, or eqn (2), if estimated through a statistical model

$$R = PVN \quad (1)$$

$$R = FVN \quad (2)$$

where:

- P is the occurrence probability, which represents the hazard event occurrence probability in the interval [0;1];
- F is the occurrence rate, which represents the observed hazard event occurrence in the interval [0;1];
- V is the vulnerability, which represents a measure of impact resistance; as it is a probability, it assumes values in the interval [0;1];
- N is the exposure, which represents the economic value of potential number of people and quantity of goods that could be involved [4].

Although the concept of risk has long been defined, in the literature there is considerable heterogeneity regarding approaches and methods of risk analysis. Risk models can be divided into two categories:

- Social risk models, that measure probabilistic (or statistic) collective damage;
- Individual risk models, that measure probabilistic (or statistic) individual damage.

They can be:

- Aggregate, when they are defined with an empirical relationship derived from global statistics about an analyzed event;
- Disaggregate, when attention is focused on each observation about an analyzed event.

In road accidents, all social risk models previously considered in eqns (1) and (2) are disaggregate [1, 3–5]. In some of these models the product between vulnerability and exposure (termed magnitude) has power with positive parameters higher than 1. Individual risk models may be either disaggregate [1, 6–11] or aggregate [8, 12]. In the disaggregate evaluation, in [1, 6, 7] risk is evaluated considering the estimation of some observed accident variables (number of exposed users and/or accidents) and in [8–11] risk is evaluated with a probabilistic function considering as input function the linear combination of variables and parameters. In particular, in [9] a Logit model is calibrated and in [8, 12] risk is specified with a negative exponential and depends on traffic flow and socioeconomic variables.

Our paper aims to overcome some limitations of models proposed elsewhere, which can be summarized by their lack of:

- completely disaggregate models, since models proposed in the literature either quantify risk in an aggregate way or, even when they are disaggregate, calculate probability and vulnerability in an aggregate way;
- risk evaluation for drivers and pedestrians moving at the same time on a road to identify the scenario where one is more exposed to accident risk than the other.

The model proposed herein allows individual risk to be determined for drivers and pedestrians under the hypothesis that a road accident has occurred. Type-of-accident probability, driver vulnerability, and pedestrian vulnerability models are calibrated. These models have a Multinomial Logit model structure, assumed as a descriptive model, and are completely disaggregate. Moreover, we construct an aggregate model that provides the probability of the pedestrian(s) involved having a specific age and sex. The system of proposed models must be considered in its totality and no model must be considered in any way separately.

Contexts can be defined through some external data that characterize the accident scenario. In practical terms, the model could support the decision planning process, allowing both comparison of hazard levels of various hypothetical scenarios and analysis of the weight and elasticity of each attribute characterizing the accident scenario.

3 PROPOSED MODEL

The model proposed in this paper follows the procedure reported in Fig. 1. First of all, the different levels for evaluating the risk and the possible alternatives are defined (Section 3.1). For each level and set of alternatives a model is specified in terms of functional form and explanatory variables and parameters (Section 3.2). The models are calibrated (Section 3.3) to obtain the best value of parameters, and then validated (Section 3.4) through formal and informal tests.

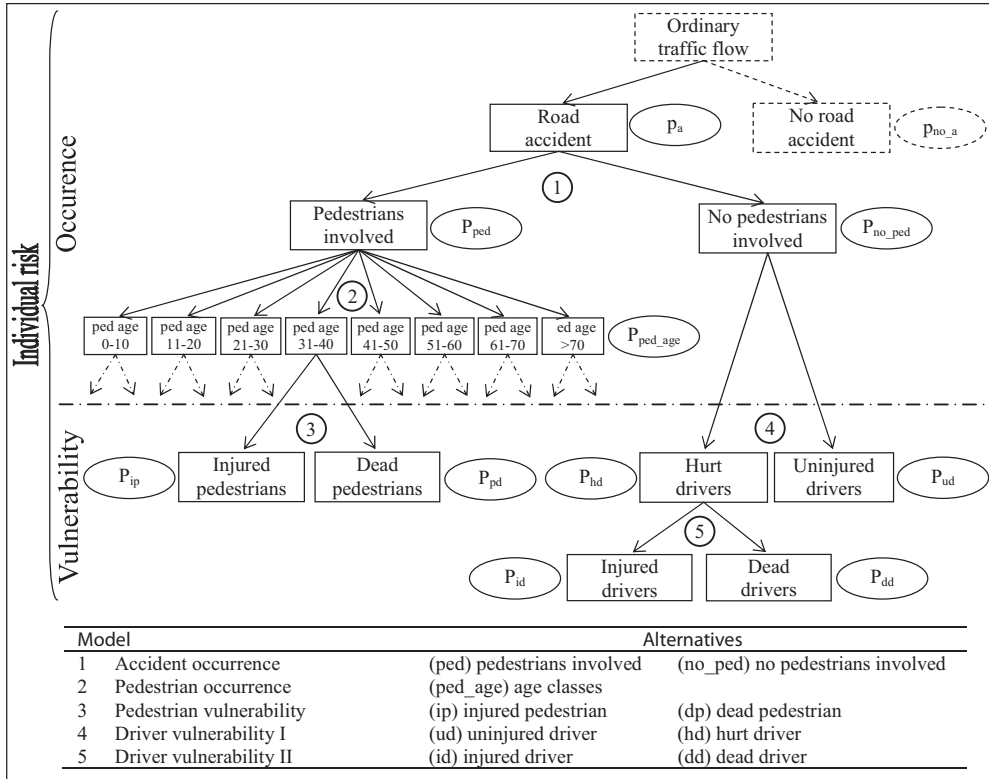


Figure 1: Tree of alternatives for a road accident.

The proposed model starts from the hypothesis that a road accident has already occurred. If one of the aims of risk analysis is also to estimate road accident probability, in the risk specification presented in this section the road accident probability, p_a , must also be introduced as reported below.

Given the notation:

- d, driver;
- p, pedestrian;
- V_d , driver vulnerability;
- V_p , pedestrian vulnerability;
- N_d , driver exposure;
- N_p , pedestrian exposure;
- no_ped, no pedestrian is involved (i.e. the accident involves two or more vehicles or an isolated vehicle).
- ped, a pedestrian is involved;
- P_{no_ped} , occurrence probability for the event no_ped;
- P_{ped} , occurrence probability for the event ped.

The two social risk functions have the following eqn (3) for drivers and eqn (4) for pedestrians:

$$R_d = p_a P_{no_ped} V_d N_d \tag{3}$$

$$R_p = p_a P_{ped} V_p N_p \tag{4}$$

The model allows us to calculate the individual risk for drivers ($p_a P_{no_ped} V_d$) and pedestrians ($p_a P_{ped} V_p$) under the hypothesis that a road accident has occurred in predefined scenarios that can be changed according to our analytical purpose.

Social risk can be obtained by multiplying the individual risk by the estimated exposure. Exposure can be estimated using demand models described and calibrated in several works in ordinary [13, 14] and emergency conditions [15, 16] implemented in various commercial software. However, the aim of the model is to develop the individual risk structure through a system of nested models. As the model is nested, it must be considered in its totality. The output is represented by the risk value for a driver and/or a pedestrian. It means that each single component of the model considered separately cannot provide any significant solution.

3.1 Alternatives

The tree of alternatives for a road accident is presented in Fig. 1. Elements above the dotted line belong to occurrence probability analysis, while elements below the dotted line belong to vulnerability analysis. A probability may be associated to each element. In the figure for each model the possible alternatives are reported.

Each alternative probability is given by the product of its upper event probabilities. For example, the individual risk of having a driver injured in an accident without pedestrian involvement may be expressed by eqn (5):

$$R_d = p_a P_{no_ped} P_{hd} P_{id} \tag{5}$$

An alternative probability represents the individual risk value for types of users and the damage incurred.

3.2 Specification

Specification consists in identifying a model's mathematical structure and attribute set. The section below describes the models adopted.

3.2.1 Multinomial Logit model specification

Accident probability and vulnerability models (models 1, 3, 4, 5) are calculated through Multinomial Logit models. They are assumed as descriptive models. Equation (6) is used to calculate the estimated frequency vector:

$$\mathbf{z} = \mathbf{u} + \boldsymbol{\varepsilon}_L \tag{6}$$

where

- \mathbf{z} is the perceived disutility or impedance vector for users relative to the available alternatives;
- $\mathbf{u} = \mathbf{u}(\mathbf{X}, \boldsymbol{\beta}_L)$ is the modeled disutility or impedance vector for users relative to the available alternatives (u_j);
- $\boldsymbol{\varepsilon}_L$ is the random residuals vector relative to the available alternatives;
- \mathbf{X} is the user attributes matrix relative to the available alternatives (\mathbf{X}_j) and attribute components;
- $\boldsymbol{\beta}_L$ is the parameters vector relative to attribute components.

The impedance u_j , for alternative j , is calculated with a linear combination of the vector of attributes \mathbf{X}_j for alternative j and the vector of parameters β_L through eqn (7):

$$u_j = \mathbf{u}(\mathbf{X}_j, \beta_L) = \beta_L^T \mathbf{X}_j \tag{7}$$

The Logit model is exclusively of the descriptive type and in no way behavioral.

In the hypothesis of Weibull distribution for the random residual, the probability associated to alternative j , $Prob(j)$, can be calculated through eqn (8):

$$Prob(j) = \exp(u_j) / \sum_k \exp(u_k) \tag{8}$$

where k is the number of alternatives.

3.2.2 Regressive model specification

The probability that the pedestrian involved is of a certain age and sex (model 2) is assumed equal to estimated accident frequency. Equation (9) is used to calculate the estimated frequency vector:

$$\mathbf{w} = \mathbf{y} + \epsilon_R \tag{9}$$

where:

- \mathbf{w} is the observed frequency vector relative to the available alternatives;
- $\mathbf{y} = \mathbf{y}(\mathbf{X}, \beta_R)$ is the model frequencies vector relative to the available alternatives (y_j);
- ϵ_R is the random residuals vector relative to the available alternatives;
- \mathbf{X} is the user attributes matrix for the available alternatives (\mathbf{X}_j) and attribute components;
- β_R is the parameter vector relative to the attribute components.

The model frequency vector associated to alternative j , y_j , is calculated with a linear combination of the vector of attributes \mathbf{X}_j for alternative j and the vector of parameters β_R through eqn (10):

$$y_j = y_j(\mathbf{X}_j, \beta_R) = \beta_R^T \mathbf{X}_j \tag{10}$$

3.3 Calibration

Calibration consists in estimating model parameters. The adopted methods are briefly presented in the sub-sections below.

3.3.1 Multinomial Logit model calibration

Accident probability and vulnerability models (models 1, 3, 4, 5) are calibrated with the Maximum Likelihood method. It supplies the values of the unknown parameters that maximize the probability of observing user choices. The latter probability is equal to the probability product of each user, i , making choice $j(i)$. Probabilities are expressed as a function of the parameter vector β_L .

The estimate of maximum likelihood is obtained by maximizing the total probability function (L). However, it is preferred to maximize its natural logarithm ($\ln L$) [14], as reported in eqn (11):

$$\beta_L^* = \arg(\max \ln L(\beta_L)) = \arg(\max \sum_i Prob(j(i))) \tag{11}$$

where $arg(max f)$ is the value of the independent variable at the point where the function f has its maximum. For the function $Prob(j)$, eqn (10) is considered.

3.3.2 Regressive model calibration

The probability that the pedestrian involved is of a certain age and sex (model 2) is calibrated with the Maximum Likelihood method. It is assumed that residual ϵ is distributed according to a normal distribution of zero average ($E[\epsilon_R] = 0$). Considering ϵ_R variance as a constant, it assumes eqn (12):

$$\beta^*_R = arg \min \sum_i (w_i - y_i)^2 \quad (12)$$

where:

- w_i is the i -th element of observed frequency vector \mathbf{w} ;
- y_i is the i -th element of model frequency vector \mathbf{y} defined in eqn (10);
- $arg(min f)$ is the value of the independent variable at the point where function f attains its minimum.

This expression coincides with that of the Least Squares method. It has the property of minimizing the sum of squared differences between observed values and theoretical values provided by the model whose parameters are unknown.

3.4 Validation

Validation consists in verifying the reasonableness and significance of parameters and ascertaining model reliability. The adopted statistical tests are presented below.

3.4.1 Multinomial Logit model validation

Accident probability and vulnerability models (models 1, 3, 4, 5) are validated by the following tests:

- informal tests on calibrated parameter signs, which verify the reasonableness of the calibrated parameter signs;
- Student's t test, which verifies the null hypothesis ($H_0: \beta^*_j = 0$) that parameter β^*_{Lj} (deriving from β^*_L) is statistically different from zero for set or model specification errors;
- the ρ^2 statistic, which measures capability to reproduce set choices [14].

3.4.2 Regressive model validation

The probability that the pedestrian involved is of a certain age and sex (model 2) is validated by the following tests:

- informal tests on calibrated parameter signs, which verify the reasonableness of the calibrated parameter signs;
- Student's t test, which verifies the null hypothesis ($H_0: \beta^*_j = 0$) that parameter β^*_{Lj} (deriving from β^*_L) is statistically different from zero for set or model specification errors;
- the ρ^2 statistic, which is a measure of estimation reliability.

4 EXPERIMENTATION

The models were constructed using a road accident database for the Italian region of Lombardy. Containing the accidents recorded in 188 municipalities in the province of Milan during the two year period from 2001 to 2002, the database consists of a series of records, each referring to a single accident for which the most important information is reported. Population data and municipal area data are provided by ISTAT [17]. Vehicle data are provided by the Italian Automobile Club (ACI) [18–20].

The available data did not allow us to calibrate the models considering numerous infrastructure explanatory variables even though the weight of these variables on accident occurrence has long been recognized. For example, Poch and Mannering [21], using seven-year accident histories from 63 intersections in Bellevue (Washington), have uncovered important interactions between geometric and traffic-related elements and accident frequencies since 1996. This means that the proposed model allows risk analysis to be performed mainly as regards the causes linked to user characteristics (i.e. sex, behavior, psycho-physical state) and environmental conditions (i.e. time band, accident site, type of vehicles involved) rather than those of a geometric type.

4.1 Alternatives and specification

Different types of attributes were tested to identify the best attribute set for model calibration. For the five calibrated models, specification was carried out by defining different impedances given by different sets of attributes to be calibrated and validated. The final attributes and specification used are those reported in Tables 1–4.

4.1.1 Accident occurrence probability model (model 1)

This model (Table 1) provides the two alternative probabilities conditioned by the occurrence of a road accident:

- *no pedestrian involved* (the accident involves two or more vehicles or an isolated vehicle);
- *pedestrian involved*.

Two impedances, u_1 and u_2 , are defined for the first and second alternative. They include driver age, number of registered vehicles/inhabitants, and some dummy variables such as the light conditions, traffic flow, and environmental factors. Therefore, in u_1 , only a shadow variable is included, while in u_2 , attributes reported in Table 1 are included.

4.1.2 Pedestrian occurrence in relation to age class (model 2)

This model (Table 2) provides the estimated pedestrian occurrence in road accidents by age class. Pedestrians are divided into age and sex categories. For each category, the observed accident frequency is estimated in relation to pedestrian age and some dummy variables depending on gender and age.

4.1.3 Pedestrian vulnerability

This model (Table 3) provides the two alternative probabilities conditional upon pedestrian involvement:

- *pedestrian injury*;
- *pedestrian fatality*.

Table 1: Accident occurrence model calibration.

Alternative	Attribute	Unit of measurement	Estimated β	t-Student
u_1	1	–	8.838	35.6
	tot_veh/pop	Number of registered vehicles/ inhabitants	4.352	16.2
	$X_{\text{day/night}}$	1 if day (7–18), 0 otherwise	0.556	13.8
	$X_{\text{rush-hour}}$	1 if accident occurs at rush- hour (7–8, 12–13, 18–20), 0 otherwise	0.205	6.4
u_2	$X_{\text{urb_road}}$	1 if urban road, 0 otherwise	0.927	9.4
	$X_{\text{built up area}}$	1 if road inside built-up area, 0 otherwise	1.383	8.6
	X_{cross}	1 if accident occurs at crossing, 0 otherwise	–0.574	–18.1
	$X_{\text{driv_age}}$	Age (years)	0.0127	11.9
		Log Initial likelihood		–42019.9
		Log Final likelihood		–15049.4
		Number of observation		60623
	ρ^2		0,64	

Model 1. Alternatives: u_1 = pedestrian involved, u_2 = no pedestrian involved.

Table 2: Model calibration of pedestrian involvement by age and gender.

Alternative	Attribute	Unit of measurement	Estimated β	t-Student
All	ped_age/100	Age (years)/100	0.595	13.4
	$(\text{ped_age}/100)^2$	$(\text{Age (years)}/100)^2$	–1.609	–9.2
	$(\text{ped_age}/100)^3$	$(\text{Age (years)}/100)^3$	1.274	7.9
	sex_m	1 if male pedestrian <10 years, 0 otherwise	0.021	1.8
	sex_f	1 if female pedestrian \geq 50 years, 0 otherwise	0.015	3.6
		Number of observations		5195
		ρ^2		0,99

Model 2. Alternatives: age classes.

Two impedances, u_1 and u_2 , are defined for the first and second alternative. They include pedestrian and vehicle attributes. In u_1 , a shadow variable is included, while in impedance u_2 , attributes related to age, speed, and class of vehicles involved are considered.

4.1.4 Driver vulnerability

This model (Table 4) is split into two levels. The first level provides the two alternative probabilities conditional upon no pedestrian involvement occurring:

- *uninjured driver*;
- *hurt driver* (injured or dead).

Table 3: Pedestrian vulnerability model calibration.

Alternative	Attribute	Unit of measurement	Estimated β	t-Student
u_1	1	–	3.526	8.2
	X_{ped_age}	Age (years)	0.02118	4.1
u_2	X_{light_veh}	1 if light vehicle (car or two-wheeled vehicle), 0 otherwise	-1.020	-4.0
	$X_{resp_speed-limit}$	1 if speed limit is observed, 0 otherwise	-0.951	-3.5
		Log Initial likelihood		-3440.1
		Log Final likelihood		-396.1
		Number of observation		4964
		ρ^2		0,88

Model 3. Alternatives: u_1 = injured pedestrian, u_2 = dead pedestrian.

Table 4: Driver vulnerability model calibration, levels I and II.

Alternative	Attribute	Unit of measurement	Model 4		Model 5	
			Estimated β	t-Student	Estimated β	t-Student
u_1	1	–	-4.883	-48.9	4.315	7.9
	X_{wheels}	1 if 4 (or plus)-wheeled vehicle, 0 otherwise	-3.365	-59.4		
	$X_{resp_speed-1}$	1 if speed limit is observed, 0 otherwise	-0.1259	-2.5	-1.0520	-3.7
u_2	X_{veh_years}	Years since vehicle registration, 0 otherwise	0.01227	3.7		
	X_{driv_age}	Years	-0.01252	-12.0	0.0099	1.3
	X_{driv_sex}	1 if male, 0 otherwise	-0.4542	-13.5	1.0010	2.5
	X_{prot_syst}	1 if driver has protection, 0 otherwise	-1.079	-21.2	-1.6050	-6.8
		Log Initial likelihood		-18396.8		-10590.6
		Log Final likelihood		-13957.9		-458.5
		Number of observations		26542		15279
		ρ^2		0.24		0.96

Model 4. Alternatives: u_1 = uninjured driver, u_2 = hurt driver. Model 5. Alternatives: u_1 = injured driver, u_2 = dead driver.

The second level provides the two alternative probabilities in the event of injury to the driver:

- *injured driver;*
- *dead driver.*

Two impedances, u_1 and u_2 , are defined for the first and second level. They include driver and vehicle attributes. In u_1 , a shadow variable is included, while in u_2 , attributes comprise driver age and vehicle characteristics.

4.2 Calibration and validation

Various calibrations are performed to reduce the objective function for each model. In the pedestrian accident probability model, each pedestrian category is replaced with its bar-center value of age (5, 15, 25, 35, 45, 55, 65, and 75 years) to apply the Least Squares method. The model has a step function trend where each step represents accident probability for a pedestrian belonging to that age category; a continuous trend (the function passes through estimated probability values) only has qualitative significance. The calibrated models are validated to verify each attribute weight and to discard attributes with the wrong parameter sign and with a low Student's t value. The best calibration and validation results are reported in Tables 1–4. In each model, all the parameters have the expected sign, confirming the validity of the models with the informal test. Considering the formal test almost all the parameters have a high significance in Student's t test. Considering also the high number of observed values, the very high mean values for the ρ^2 statistics indicate the goodness of fit from the models and the data.

Each model can also be interpreted in relation to the value of a single parameter. Just a few indications are reported in this section. Other analyses can be performed by observing and applying the models in relation to the scenario and the real application, as briefly reported in section 5 for the city of Reggio Calabria.

In the following, some examples of dependency between probability and attributes are reported. In relation to model 1, the probability that a pedestrian is involved increases during the day in built-up areas and decreases at crossings. In relation to model 2, the probability that a pedestrian is involved increases for males under 10 years and females over 50 years of age. In relation to model 3, the probability of a pedestrian dying increases at night and if the speed limit is not respected. In relation to models 4 and 5, the probability of driver fatality increases if the speed limit is not observed and he/she has no protection system (helmet, belt).

With regard to its application, the model could be used to support the decision planning process [22–24]. For example, let us suppose that the network administrator wants to analyze risk for a road, considering a user attribute like pedestrian age. Accident scenarios change with pedestrian age while the other model attributes are kept fixed. Individual risk might prove acceptable for adult pedestrians. However, if the road is close to a school, with a large flow of young pedestrians crossing it every day, then individual risk might prove unacceptable and an intervention must be planned. By the same token it is assumed that the analysis is performed with a road service attribute like regulation type. Individual risk for drivers might prove unacceptable if the road is regulated as extra-urban but acceptable in the case of urban regulation.

In the literature various methods and models to plan and assess interventions to improve traffic safety have been presented. Delhayé [25] sought to model traffic safety using a cost of driving function depending on speed and value of time. Al-Kaisy [26] investigated the effectiveness of signing at intersections in reducing conflict between pedestrians and turning vehicles.

5 TEST IN A REAL CONTEXT

The test was performed in the city of Reggio Calabria (Italy) for the following purposes: to verify model transferability and reliability when applied to a different context from that on which the model was calibrated, to analyze and plot probability and vulnerability curves, and analyze and plot individual risk curves as regards predefined accident scenarios.

Since the proposed model considers user characteristics and environmental conditions rather than geometric variables, it might not be effective in the same way if applied abroad due to economic, social, and infrastructural differences with the other countries and different road user behavior.

For our test in Reggio Calabria we used the 2005 road accident database, consisting of a series of 1,146 records in all. Each record refers to a single accident and contains information on vehicles, drivers, pedestrians, and environmental conditions. Model transferability and reliability were verified using quantitative statistics. In particular, the Kolmogorov–Smirnov test was used to verify that the model, calibrated on the Milan road accident data set, fits the data of its counterpart in Reggio Calabria. This test was applied to continuous distributions, comparing theoretical distribution with observed distribution.

Considering an observable or a measurable event like a road accident, let:

- $F_o(x)$, be the observed cumulative distribution for the observations;
- $F_t(x)$, the theoretical cumulative distribution.

The Kolmogorov–Smirnov test is given by eqn (13):

$$D = \max_x (F_o(x) - F_t(x)) \quad (13)$$

The following hypotheses with a pre-arranged significance level are to be verified:

- H_0 hypothesis, the sample is distributed like $F_t(x)$;
- H_1 hypothesis, the H_0 hypothesis is refused.

Acceptance of either one hypothesis or the other is determined by comparing D with the critical values D_c , known for different levels of significance. If D is smaller than D_c , the hypothesis H_0 is accepted; otherwise the hypothesis H_1 is accepted.

In the Reggio Calabria data, considering the probability function reported in this paper and calibrated with Milan data, the Kolmogorov–Smirnov test was verified by each model for significance levels of 90% and 95%, which indicates good model reliability. The test is carried out for eight scenarios to analyze the model obtained in relation to each level and global individual risk for Reggio Calabria.

In relation to the accident occurrence model (model 1), the purpose is to point out that probabilities of pedestrian involvement and no pedestrian involvement change with (Table 5):

- driver age and time slice (scenarios I and II);
- location (scenarios I, III, IV).

The curves obtained are reported in Fig. 2. The results indicate that the probability of pedestrians being involved is higher during the day and at rush-hour; hence the

no-pedestrian-involved probability is higher at night and not at rush-hour. Moreover, as regards location, probability of pedestrian involvement is higher in straight road while the no-pedestrian-involved probability is higher outside built-up areas. This may be explained simply: pedestrians move by day and at rush-hour, mainly with an accident concentration in this time slice. Hence, most accidents at night are statistically those where no pedestrians are involved. Moreover, the probability of pedestrian involvement is lower at crossings because the presence of traffic lights and/or signs makes crossing safer. There is a higher probability of pedestrians not being involved outside built-up areas where traffic is not congested and speeds are higher.

In relation to the vulnerability models (models 3 and 4), the purpose was to show how injured pedestrian and uninjured driver probabilities change with (Table 5):

- pedestrian and driver age (scenarios V and VI);
- type of vehicle (scenarios VII and VIII).

Table 5: Scenarios tested for the calibrated models.

Scenario	Time band	Rush-hour	Road type	Location	Position
I	day	yes	urban	inside built-up area	straight
II	night	no	urban	inside built-up area	straight
III	day	yes	urban	inside built-up area	crossing
IV	day	yes	urban	outside built-up area	straight

Scenario	Type of vehicle	Speed limit observed
V	light vehicle (car; two-wheeler)	Yes
VI	heavy vehicle	Yes

Scenario	Type of vehicle	Speed limit observed	Vehicle age	Driver sex	Protection systems
VII	4 (or +) wheels	Yes	3	male	no
VIII	2 wheels	Yes	3	male	no

Scenario	Type of vehicle	Effects on pedestrian
A	heavy vehicle	Injured
B	light vehicle	Injured

Scenario	Time band	Rush hour	Type of vehicle	Effects on driver
C	night	No	two- wheeler	injured or dead
D	day	Yes	four (or +) wheeler	injured or dead

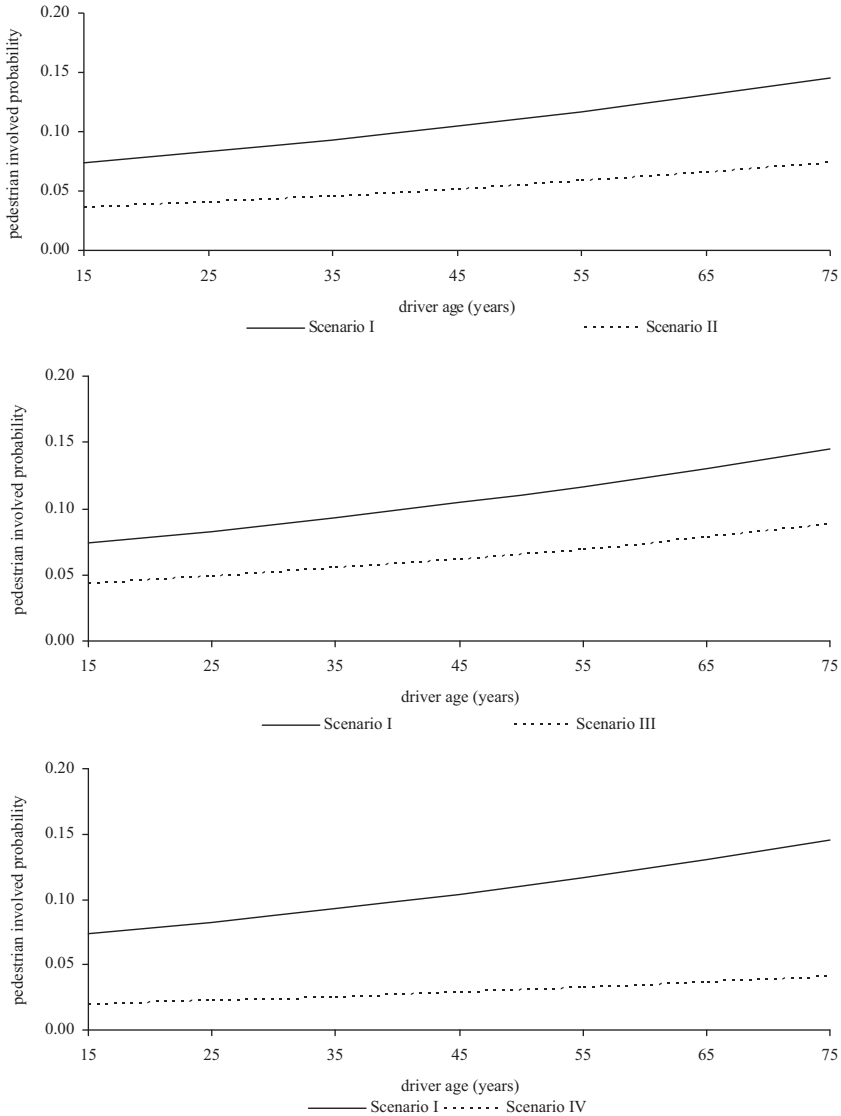


Figure 2: Probability of pedestrian involvement measured against driver age in the proposed scenarios (model 1, accident occurrence probability) conditional upon a road accident.

The curves obtained are reported in Figs 3 and 4. The results show that the probability of pedestrian fatality is higher for heavy vehicles and that probability of drivers being uninjured is lower for two-wheeled vehicles.

Risk curves for drivers and pedestrians are plotted for the more and less dangerous scenarios. The results obtained are reported in Figs 5 and 6. The proposed model allows hazard levels to be compared in various hypothetical situations. It is shown that the risk of pedestrian injury is almost the same whether he/she is hit by a light or heavy vehicle (Fig. 5). This means that the consequences of pedestrian accidents are linked much more closely to the absence of

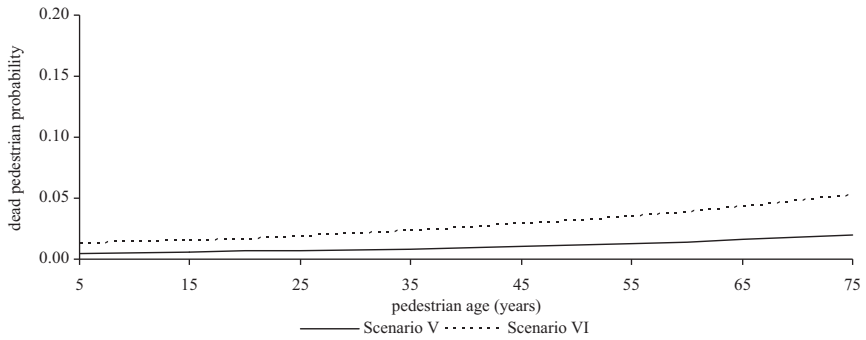


Figure 3: Probability of pedestrian fatality measured against pedestrian age in the proposed scenarios (model 3, pedestrian vulnerability model) conditional upon a road accident with a pedestrian involved.



Figure 4: Uninjured driver probabilities measured against driver age in the proposed scenarios (model 4, driver vulnerability model) conditional upon a road accident.

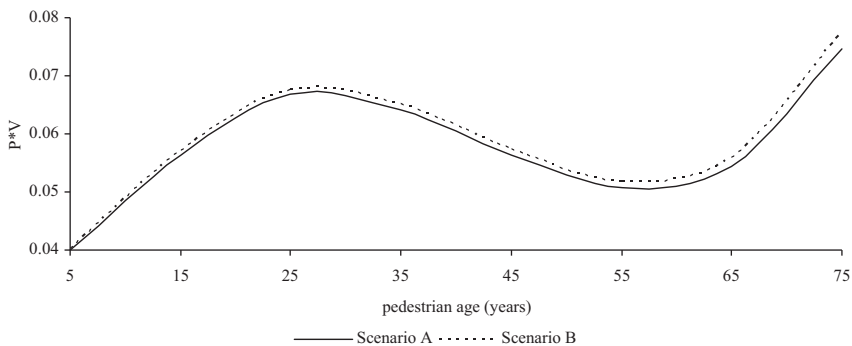


Figure 5: Comparison between pedestrian individual risk functions upon changes in the proposed scenario.

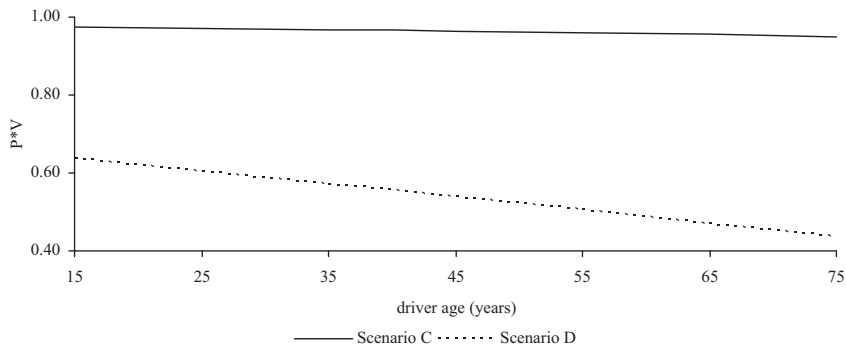


Figure 6: Comparison between driver individual risk functions upon changes in the proposed scenario.

any protection. By contrast, the driver's risk of injury is higher in the case of accidents on two-wheeled vehicles than on four (or more)-wheeled vehicles (Fig. 6).

Starting from these surveys, it is possible to define types of intervention to be pursued: in the first case it is advisable to adopt preventive measures to reduce accident probability (such as seeking to separate pedestrian flows from vehicle flows), while in the second case mitigation measures should be adopted to improve two-wheeled vehicle safety, as accident probability may be considered constant during planning.

Obviously the most advisable interventions to plan must be identified in terms of estimated exposure value, which is linked to the economic and social characteristics of the road accident environment (i.e. urban area, extra-urban area, residential area, industrial area, presence of offices and/or schools, etc.), and the composition and volumes of vehicle and pedestrian flows.

For example, as regards Fig. 5, to reduce pedestrian risk different interventions may be planned which, under increasing pedestrian flows, can vary from changes to road signs (e.g. introduction of larger signals or lighting signs), to the introduction of traffic lights, the construction of a pedestrian underpass or overpass, the introduction of barriers to prevent crossing, or the creation of pedestrian zones. Instead, as regards Fig. 6, to reduce 2-wheeled vehicle driver risk, regulatory interventions may be planned such as introducing mandatory protective wear in relation to the power of the vehicle driven and minimum driver experience (years).

6 CONCLUSIONS

In this paper we presented an individual risk model in the case of a road accident. The purpose was to estimate the probability of a pedestrian being involved (besides the driver) and the probability of their being injured under pre-assigned conditions. The model also allows the weight and elasticity of each attribute characterizing the accident scenario to be analyzed.

Although the definition of risk has long been established, there is substantial heterogeneity in approaches and methods for risk analysis. Therefore, this paper presented a risk model whose formal expression is derived directly from risk definition. The type of accident probability, driver vulnerability, and pedestrian vulnerability models are calibrated. Such models have a Multinomial Logit model structure assumed as a descriptive model. Importantly, the models are of a completely disaggregate type.

Pedestrian accident probability was also investigated in depth. An aggregate model was constructed that provides the probability of the pedestrian involved being of a certain age and sex. Finally, a test was carried out on the city of Reggio Calabria (Italy) to verify model transferability and reliability using quantitative statistics. In particular, the Kolmogorov-Smirnov test was used to verify that the model, calibrated on one data set, fits the data of another data set. Model results were validated using accident data recorded in Reggio Calabria in 2005 and the accuracy of this test was verified by each model for significance levels of 90% and 95%. Probability, vulnerability, and individual risk curves were plotted for some accident scenarios.

The methodology proposed gave good results with the data used. To ensure that the models have a general application, they had to be specified, calibrated, and validated in other contexts and in the same context in different time periods. However, it was pointed out that since the proposed model considers user characteristics and environmental conditions rather than geometric variables, it follows that it may not be effective in the same way if applied abroad due to economic, social, and infrastructural differences with other countries and different road user behavior. This means that, despite the general nature of its theoretical framework, if the model has to be applied outside Italy, it must be specified, calibrated, and validated once again to identify the new set of the most significant attributes for the context in question.

There is considerable room for improvement in many aspects of the model. While it currently concerns accident prediction, the model is time-independent. Future research efforts will focus on specifying a dynamic model even if this entails consideration of extensive spatial and temporal data and the model's analytical structure could become very complex. Many explanatory variables will have to be included (i.e. road characteristics) in model specification to improve risk analysis, given the role played by infrastructure characteristics on the occurrence of road accidents. Finally, pre- and post-analysis will have to be carried out to verify the results before and after interventions.

REFERENCES

- [1] Carpignano, A., L'analisi di rischio. Duplicated lecture notes of the Master "Affidabilità sicurezza e manutenzione dei sistemi tecnologici complessi". Politecnico di Torino, 2002.
- [2] ISTAT (Italian Institute of Statistics), *Incidenti stradali – anno 2009*. www.istat.it.
- [3] Center for Chemical Process Safety, *Guidelines for Chemical Transportation Risk Analysis*, American Institute of Chemical Engineers: New York, 1995.
- [4] Russo, F. & Vitetta, A. (eds), Risk evaluation in a transportation system. *International Journal of Sustainable Development and Planning*, **1(2)**, pp. 170–191, 2006. [doi:http://dx.doi.org/10.2495/SDP-V1-N2-170-191](http://dx.doi.org/10.2495/SDP-V1-N2-170-191)
- [5] Evans, A.W., Estimating transport fatality risk from accident data. *Accident Analysis*, **35**, pp. 459–472, 2003. [doi:http://dx.doi.org/10.1016/S0001-4575\(02\)00024-6](http://dx.doi.org/10.1016/S0001-4575(02)00024-6)
- [6] Keall, M.D., Pedestrian exposure to risk of road accident in New Zealand. *Accident Analysis*, **27**, pp. 729–740, 1995. [doi:http://dx.doi.org/10.1016/0001-4575\(95\)00019-V](http://dx.doi.org/10.1016/0001-4575(95)00019-V)
- [7] Lee, C. & Abdel-Aty, M., Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis*, **37**, pp. 775–786, 2005. [doi:http://dx.doi.org/10.1016/j.aap.2005.03.019](http://dx.doi.org/10.1016/j.aap.2005.03.019)
- [8] Delfino, G., Rindone, C., Russo, F. & Vitetta, A., Risk analysis in road safety: a model and an experimental application for pedestrians. *Proc. of the European Transport Conference*, Strasbourg, France, 2005.

- [9] Mannering, F.L. & Grodsky, L.L., Statistical analysis of motorcyclists' perceived accident risk. *Accident Analysis*, **27**(1), pp. 21–31, 1995. doi:[http://dx.doi.org/10.1016/0001-4575\(94\)00041-J](http://dx.doi.org/10.1016/0001-4575(94)00041-J)
- [10] Kleinbaum, D.G., Kupper, L.L. & Morgenstern, H., *Epidemiologic Research*, Van Nostrand Reinhold: New York, 1992.
- [11] Rodgers, G.B., Bicyclist deaths and fatality risk patterns. *Accident Analysis*, **27**(2), pp. 215–223, 1995. doi:[http://dx.doi.org/10.1016/0001-4575\(94\)00063-R](http://dx.doi.org/10.1016/0001-4575(94)00063-R)
- [12] Evans, L., *Safety and the Driver*, Van Nostrand Reinhold: New York, 1991.
- [13] Cascetta, E., *Transportation Systems Engineering: Theory and Methods*, Kluwer: Dordrecht, The Netherlands, 2001.
- [14] Domencich, T.A. & Mc Fadden, D., *Urban Travel Demand: A Behavioural Analysis*, Elsevier: New York, 1975.
- [15] Vitetta, A., Musolino, G. & Marciandò, F.A., Safety of users in road evacuation: calibration of cost functions and simulation. *WIT Transactions on the Built Environment*, **101**, pp. 715–725, 2008. doi:<http://dx.doi.org/10.2495/UT080691>
- [16] Russo, F. & Chila, G., Safety of users in road evacuation: demand models. *WIT Transactions on the Built Environment*, **96**, pp. 773–782, 2007. doi:<http://dx.doi.org/10.2495/UT070731>
- [17] ISTAT (Italian Institute of Statistics), *Censimento 2001 – Censimento popolazione ed abitazioni* (www.istat.it), 2005.
- [18] A.C.I (Automobil Club Italia), *Autoritratto 2001* (www.aci.it), 2002.
- [19] A.C.I (Automobil Club Italia), *Autoritratto 2002* (www.aci.it), 2003.
- [20] A.C.I (Automobil Club Italia), *Autoritratto 2004* (www.aci.it), 2005.
- [21] Poch, M. & Mannering, F., Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, **122**(2), pp. 105–113, 1996. doi:[http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(1996\)122:2\(105\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105))
- [22] Russo, F. & Rindone, C., Safety of users in road evacuation: modelling and DSS for LFA in the planning process. *WIT Transactions on Ecology and the Environment*, **120**, pp. 453–464, 2009. doi:<http://dx.doi.org/10.2495/SDP090421>
- [23] Russo, F. & Rindone, C., Safety of users in road evacuation: planning internal processes and guidelines. *WIT Transactions on the Built Environment*, **96**, pp. 825–834, 2007. doi:<http://dx.doi.org/10.2495/UT070781>
- [24] Russo, F. & Vitetta, A., A topological method to choose optimal solutions after solving the multi-criteria urban road network design problem. *Transportation*, **33**(4), pp. 347–370, 2006. doi:<http://dx.doi.org/10.1007/s11116-005-3507-7>
- [25] Delhayé, E., Traffic safety: speed limits, strict liability and a km tax. *Transportation Research Part A*, **40**(3), pp. 205–226, 2006. doi:<http://dx.doi.org/10.1016/j.trra.2005.06.003>
- [26] Al-Kaisy, H.N.A., Investigation of signing to reduce conflicts between pedestrians and turning vehicles at signalized intersections. *Transportation Research Part A*, **30**(1), pp. 64–65, 1996. doi:[http://dx.doi.org/10.1016/0965-8564\(96\)81129-3](http://dx.doi.org/10.1016/0965-8564(96)81129-3)