Journal of Transportation Safety & Security

ISSN: 1943-9962 (Print) 1943-9970 (Online) Journal homepage: http://www.tandfonline.com/loi/utss20

Categorical Modeling to Evaluate Road Safety at the Planning Level

Sara Ferreira & António Couto

To cite this article: Sara Ferreira & António Couto (2012) Categorical Modeling to Evaluate Road Safety at the Planning Level, Journal of Transportation Safety & Security, 4:4, 308-322, DOI: 10.1080/19439962.2012.679385

To link to this article: https://doi.org/10.1080/19439962.2012.679385

Accepted author version posted online: 30 Mar 2012.
Published online: 30 Mar 2012.

Journal of Transportation Safety & Security, 4:308–322, 2012

ISSN: 1943-9962 print / 1943-9970 online
DOI: 10.1080/19439962.2012.679385

Categorical Modeling to Evaluate Road Safety at the Planning Level

SARA FERREIRA AND ANTÓNIO COUTO

Civil Engineering Department, School of Engineering, Porto University, Porto, Portugal

The most efficient strategy to ensure long-term road network safety is to integrate safety analysis into the planning process of a network or a corridor. Safety planning decision-support tool outcomes should be reliable and realistic, taking into account the main characteristics of this particular level, which is characterized by scant and generalized data. However, the tools developed and presented in previous studies are based on models with a quantitative response. To develop a more suitable tool while maintaining a measure assessment character, this work presents a qualitative response model whose outcome is the risk of occurring three degrees of hazards: low, medium, and high. In this study, an ordered probit model was applied to an urban road network using Porto city data. Hazard categories were defined using accident frequency to reflect a measure of the safety of the road network studied. The developed model provides a safety risk analysis considering road data that are easy to gather or estimate at the planning level. In addition, an analysis of hypothetical scenarios for two sample segments is presented to illustrate an application of the categorical model in typifying an accident risk analysis at a planning level.

Keywords ordered probit, road safety, transportation planning, risk

1. Introduction

Around 85% of the European Union's gross domestic product (GDP) is generated in cities (Commision of the European Communities, 2009). Efficient transport systems are needed to support their economy and the welfare of their inhabitants. Currently, urban mobility policies face a challenge in providing sustainable transport due to different principles that sometimes oppose each other. Research priorities and agendas for urban mobility are focused mainly on energy and the environment. However, road safety is a social concern that should be included especially in long-term land use planning and access management decisions. Safety can also be one of the criteria used to evaluate and prioritize projects submitted for funding. Planners may encourage the consideration of safety improvements to all projects (AECOM, Bellomo-Mcgee Inc., and Ned Levine & Associates, 2002; Washington, et al., 2006).

Over the last three decades, the number of road accidents in Portugal has decreased by 74%. Despite this, the Portuguese average continues to be higher than the European

Address correspondence to Sara Ferreira, Civil Engineering Department, School of Engineering, Porto University, Rua Dr. Roberto Frias, Porto 4200-465, Portugal. E-mail: sara@fe.up.pt

average in terms of the number of fatalities per million inhabitants. A total of 47% of these fatalities and 71% of accidents resulting in injury occur on urban Portuguese roads (Autoridade Nacional de Segurança Rodoviária [ANSR], 2009). In this context, road safety considerations should be explicitly included and weighed at an urban planning level. In fact, ideally, road accidents should be prevented by anticipating the risk of accident occurrence when the road network is being planned. In addition, alternative network options should be recommended.

To provide safety planning decision-support tools, several studies have presented accident prediction models. The majority of these models are area-level models (macrolevel), usually based on data aggregated at the traffic analysis zone level (TAZ) (Guevara et al., 2004; Hadayeghi et al., 2003, 2010; Levine et al., 1995; Naderan & Shahi, 2010; Sayed & Lovegrove, 2006). In addition, some models applied to road-level aggregation have also been developed for evaluating alternative road networks (Ferreira & Couto, 2011; Lord & Persaud, 2004; Tarko et al., 2008).

For the above-mentioned accident prediction models developed for the planning level, which are featured by a lack of road network information, the outcome is the number of accidents per year, that is, a quantitative response. However, in the authors' opinion, this response is not consistent with the poor level of road network characterization available at the planning level. Moreover, at this level, the goal is to select the best solution among different scenarios, taking into account a safety indicator that evaluates safety in a broader manner instead of using a specific number of accidents that is dependent on many factors that are unknown at this level.

In this sense, this study presents a qualitative response model whose outcome is a qualitative measure that characterizes the road network scenario by the degree of hazards. The hazard categories were defined by a range of accident numbers to reflect the magnitude of the safety of the road network studied, independent of the characteristics of the road entity, thus providing a more appropriate tool while maintaining the measure assessment character. Three categories illustrating the degree of hazard based on a range of numbers of accidents were defined as low, medium, and high. Using this qualitative approach, road entity scenarios can be assessed by the occurrence probability of the three degrees, thus providing a risk analysis that offers selection support for a safe solution. To achieve this, an ordered probit model (OPM) was developed and applied to an urban road network using Porto (Portugal) data covering a 5-year period. The independent variables used in the OPM geometrically and functionally characterize segments based on data gathered and/or estimated at the planning level: traffic volume, segment length, number of minor intersections (intersections with minor roads, usually without associated traffic data such as access roads), land use, and road function classification. Although the OPM was developed for urban segments to provide a comparison between several road network scenarios, the conceptual development can be applied to an aggregate model.

The results obtained using the parameter estimates illustrate the impact of the independent variables on the probability of the three categories. Moreover, these results in agreement with those of other studies using these variables in count-data models (Greibe, 2003; Ivan et al., 2000; Karlaftis & Tarko, 1997; Mountain et al., 1996, 1998; Wedagama et al., 2006; Wier et al., 2009) as well as with the results obtained from a count-data model applied to the data set used in this study (Ferreira, 2010; Ferreira & Couto, 2011). Furthermore, hypothetical segment scenarios were set to demonstrate a possible planning exercise based on risk analysis as well as the consequences of changing the exogenous factors.

The rest of this article presents the background, data, and methods; estimation results; analysis of hypothetical scenarios and a summary; and conclusions regarding the development and application of an ordered response model as a safety planning model, in that order.

2. Background

Several studies have been presented to overcome a lack of available planning-level tools as referred by De Leur and Sayed (2002). From these studies two different levels of data aggregation have emerged: road level and area level. The latter are usually based on data aggregated at the TAZ and make up the exposure and road network data represented by vehicle kilometers traveled, network density, population, number of employees, and so on. However, Tarko et al. (2008) pointed out that area-level models are useful for evaluating transportation and safety-related policies and area-wide solutions but less practical for evaluating specific road improvements, screening networks for dangerous roads, estimating the impact of road characteristics, or merely predicting future accidents in specific road entities. In this sense, road-level models have been developed by some authors (Ferreira & Couto, 2011; Lord & Persaud, 2004; Tarko et al., 2008). In this case, the road network was characterized by a series of nodes/intersections and links/segments, and therefore separate models were developed using variables such as traffic volume (usually defined as the annual average daily traffic [AADT]), number of lanes, segment length, number of intersection arms, and so on.

Different approaches for both types of data aggregate-level of accident prediction models have been presented being however, noteworthy that a common technique, namely, generalized linear modeling (GLM) with the assumption of negative binomial (NB) error distribution has been widely used (Guevara et al., 2004; Hadayeghi et al., 2003; Lord & Persaud, 2004; Naderan & Shahi, 2010; Sayed & Lovegrove, 2006; Tarko et al., 2008). The GLM procedure, normally used in the above-mentioned models, comprises the estimation of parameters to represent the average relationship between the dependent variable, typically number of accidents per TAZ or per intersection and per segment depending on the level of data aggregation, and each explanatory variable. Hence, these are count-data models whose outcome is a quantitative measure of safety.

At the road-planning level, especially in the decision-making process, an assessment of safety by a quantitative measure is not suitable when a poor level of information of road network characteristics is available. In fact, at this level, a safety measure is needed to evaluate and compare alternative scenarios but not necessarily to predict a number of accidents whereas this can be done later applying existing models and with more accurate results.

In this sense, an alternative approach to evaluate safety at the road planning and decision level is proposed in this work. This approach is based on a categorical modeling that defines the dependent variable as an indicator of a discrete choice, that is, the dependent variables are merely a coding for some qualitative outcome (Greene, 2008). In this approach, a general framework of probability models are used to link the outcome to a set of factors (Greene, 2008):

$$\text{Prob(event j occurs)} = \text{Prob}(\text{Y}- = \text{j}) = \text{F[effects, parameters]} \tag{1}$$

where the "event" is an individual's choice among a set of alternatives.

Different types of categorical modeling techniques have been widely applied in economics and modeling of transportation behavior. In the field of road safety, categorical modeling techniques have been more recently used to model accident severity whereas the severity level, such as no injury, injury and fatality, is a discrete outcome. The most common techniques applied to analyze accident severity were the multinomial logit, nested logit, and ordered probit and logit formulation (Abdel-Aty, 2003; Carson & Mannering, 2001; Eluru et al., 2008; Kockelman & Kweon, 2002; Savolainen et al. 2011; Wang & Abdel-Aty, 2008). These models can be grouped in two response mechanisms: the ordered response (ordered probit and ordered logit) and unordered response (multinomial logit, nested logit, and multinomial probit). The ordered response mechanism has the advantage of being parsimonious in structure because it imposes the restriction that the regression parameters are the same for different severity levels. Hence, the adjacent severity levels are correlated. On the other hand, the unordered response mechanism is based on a utility-maximization principle hypothesis and thus the severity levels are not presumed to correspond to the successive partition of a uni-dimension latent variable (Bhat & Pulugurta, 1997).

There is not a clear consensus in the choice of the response mechanism to apply in the accident severity analysis. Abdel-Aty (2003) compared the multinomial logit, nested logit, and OPMs for driver's injury severity at toll plazas and concluded that the nested logit model produced the best fit. However, other authors (Kockelman & Kweon, 2002; Train, 2003) pointed out that such specification does not actually fit the structure of the ordinal data.

In the context of accident frequency, the discrete choice models were seldom used taking into account the "count" nature of frequency data. Qi et al., (2007) applied a random effects OPM to model and predict accident likelihood (online model) taking into account the preponderance of nonaccident and few accident cases in a short time period analyzed. In this study three responses (choices) were considered: nonaccident (0), one accident (1), and more than one accident (>1) per time interval. Because the responses were ordinal, ordered response model were used. The study results illustrate that the model performs well in identifying factors associated with road accidents and in forecasting the likelihood of accidents based on time-varying and site-specific parameters.

Based on the review above, it is clear that the applications of discrete choice models were almost limited to the analysis of accident severity. Besides, on the authors' knowledge, nobody has so far examined the possibility of applying a discrete choice model to evaluate safety of different road solutions at the road-planning level.

3. Data and Method

To illustrate the application of a discrete choice model for assessing safety at the road-planning level using a qualitative measure as the outcome, an urban segment model was developed using data from Porto, Portugal. At the planning level, segment models can be used to analyze scenarios for a corridor or of a road network entity (an intersection model was also developed in a previous work [Ferreira, 2010] to provide a complete road network analysis). Nonetheless, the conceptual approach presented in this article can be applied to a data aggregate model like such as an area-level model, whereas the results suggest considerable potential for further applications of these types of models.

The data used in this study consist of accident data from urban segments classified by local and principal distributor roads collected over a 5-year period (1 January 2001 to 31 December 2005). Accident data were obtained from the official Portuguese Police Security database, covering all police-recorded accidents with local information (accidents resulting in injury and accidents resulting in property damage only). All accidents were related to their specific location by applying a geographic information system (GIS) tool. The data consist of 5,650 police-recorded accidents; of these, 1,183 were personal injury accidents and 4,467 were property-damage-only accidents that were related to 396 segments.

Using these injury accidents and property-damage-only accidents, hazard categories were set to reflect the magnitude of the hazard from an accident frequency standpoint (an accident severity standpoint could be used as well) of the road network studied. Taking into account the road-planning context, instead of a risk factor given by the accident frequency per unit of exposure (i.e., a measure of "accident opportunity"), usually the accident frequency per length or traffic volume unit, the categories were based exclusively on accident frequency. In a planning context, when analyzing safety, one is more focused on choosing between alternative road options or assessing the factor exposure impacts, followed by a risk factor evaluation.[1] In the proposed approach, the risk of different road alternatives is analyzed after the probabilities of each category are determined.

Three categories[2] illustrating the degree of hazard by a range of numbers of accidents were defined: low, medium, and high. The first category represents the 0 to 2 accidents. Such low accident numbers are less likely to be related to safe characteristics of the road entity, but rather to the unusual driver behavior, for example. The second category aims to reflect a substantial range of accident numbers that ultimately result from the unsafe features of an urban segment. The latter intends to illustrate the high range of accident frequencies that represent an unacceptable safety situation. The boundary between the last two categories can be defined by a flexible criterion that can be easily determined when applied in a jurisdiction different from the one developed. In this study, the boundary was considered to be the accident number given by the 90th percentile of all recorded observations, which is eight accidents. This number ensures that a serious situation as defined in terms of safety is an improbable situation. Therefore, the three responses used to model the data were: low (0–2 accidents), medium (3–8 accidents), high (>8 accidents). The sample distribution for these responses is: low (64%), medium (28%), high (8%).

These responses were related to road network attributes represented by independent variables.

The variables selected as independent variables were chosen by considering information that can be gathered and/or estimated at the road-planning level. Therefore, to geometrically characterize a segment, the two most common variables besides the traffic flow were used: the segment length[3] and the number of minor intersections per segment length. These variables are usually used to homogenize segments and are easy to determine at the planning level by, for example, using a GIS tool. Furthermore, to describe the urban environment, road design and flow pattern characterizing a road network, land use, and road function classification variables were included. These variables, especially the land use variable, have been extensively studied as independent variables (Dissanayake et al., 2009; Greibe, 2003; Ivan et al., 2000; Wedagama et al., 2006). In fact, the main decisions required in the urban planning process are related to land use and road function classification.

In this study, five different types of land use based on the municipal master plan were taken into consideration as dummy variables: Land Use 1 (LU1)—high density of buildings; Land Use 2 (LU2)—low density of buildings; Land Use 3 (LU3)—industrial area; Land Use 4 (LU4)—community building area (educational buildings and sports grounds); Land Use 5 (LU5)—historic center area. Figure 1 illustrates the land use classification of Porto city.

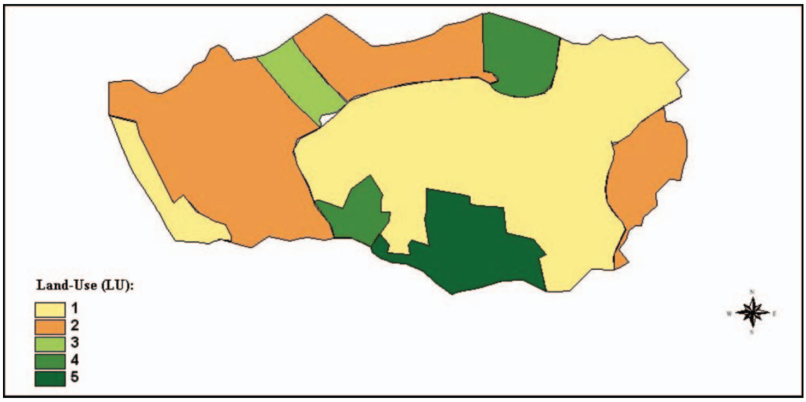

Figure 1. Land use classification of Porto city. 1: Area with high density of buildings; 2: area with low density of buildings; 3: industrial area; 4: community building area; 5: historic center area. (Color figure available online.)

In addition, the municipal master plan defines four road classes, namely, arterial roads, principal distributor roads, local distributor roads, and access roads. However, only principal distributor roads and local distributor roads were used (also as dummy variables) due to the fact that arterial roads have characteristics similar to those of a highway and that there is a lack of traffic flow data for access roads. Figure 2 shows the road function classification of the Porto road network. A time trend variable was also included in this study to reflect a potential change in the overall accident level over time. A number of studies have included a time trend variable that allows for changes in terms of risk over time (Fridstrom et al., 1995; Greibe, 2003; Mountain et al., 1998). It has been previously demonstrated that this variable has a negative effect on the frequency of accidents, thus leading to an overall improvement in relation to safety education, enforcement, vehicles, and so on. In this model, a geometric form was used, assuming that the "general safety development" is the same from year to year (Mountain et al., 1998). To test for possible temporal correlations, a likelihood ratio

Figure 2. Road function classification of the Porto road network. Local Dist: local distribution roads; PrincDist: principle distribution roads; Arterial: arterial roads; Access: local access roads. (Color figure available online.)

test was conducted as described by Poch and Mannering (1996), revealing that this issue does not significantly affect the resulting estimates.

Finally, because traffic flow values are not available for all road networks, the AADT was estimated by the Porto simulation and assignment of traffic to urban roads (SATURN) traffic model and data provided by permanent counting stations located throughout the principal city zones belonging to the Urban Traffic Center were used. The SATURN model was developed in 2005 for the peak hourly flow (PHF) of the Porto road network and involved all the recommend development stages (specification, coding, error checking, calibration, validation) (Tavares, 2003). The PHF was expanded to AADT every 5 years by applying zone factors that relate the PHF with the daily traffic flow calculated using data provided by permanent counting stations located throughout the principal city zones belonging to the Urban Traffic Center. Traffic simulations have been used in other studies related to the planning level (Hadayeghi et al., 2010; Lord & Persaud, 2004), and it was noted by Lord and Persaud (2004) that the accuracy of such predictions is directly related to the precision of the traffic flow estimates. Table 1 presents a statistical description of the dependent and independent variables used in the frequency models.[4]

As is shown in Table 1, a high degree of heterogeneity is present in some of the variables, namely, the density of minor intersections and the segment length, reflecting the urban environment. The former has a maximum value that implies an average distance between minor intersections of about 30 meters. However, only about 2% of the segments have a minor intersection density greater than 20 per kilometer. Additionally, low segment length values (fewer than 50 meters) are exhibited by only 11% of the segments studied. These extreme value observations were maintained in the data set to describe all of the urban road networks.

Because the three responses (choices) mentioned above are ordinal and have an eventual correlation between adjacent categories, ordered response models were selected. Ordered response probit and logit models were tested. As expected, the results obtained were very similar because these ordinal model forms are essentially equivalent and differ only in whether a logistic or a normal distribution is used for the stochastic component in the

Table 1

Statistical description of variables used in accident frequency models (5-year)

Variable	Min.	Max.	Average	*SD*
Accident frequency	0	27	2.85	3.90
AADT	142.37	64,067.80	15,240.38	11,699.49
Segment length (in meters)	20.71	3342.78	313.18	352.56
Number of minor intersections per kilometer	0.00	32.04	4.57	5.18
High density of buildings (LU1)	0	1	0.55	0.50
Low density of buildings (LU2)	0	1	0.22	0.42
Industrial (LU3)	0	1	0.03	0.17
Community buildings (LU4)	0	1	0.06	0.23
Historic center (LU5)	0	1	0.14	0.35
Local distributor roads	0	1	0.50	0.50
Principal distributor roads	0	1	0.50	0.50

AADT = annual average daily traffic.

latent propensity that is assumed to underlie the observed accident frequency. Because the ordered response probit model (OPM) results were slightly better, this form was selected.

This model is specified based on a latent regression model as illustrated below:

$$y_n^* = x_n'\beta + \varepsilon_n \quad \mathrm{n} = 1, \ldots, \mathrm{N} \tag{2}$$

where,

y_n^* unobserved components
x_n' vector of independent variables
β vector of parameters
ε_n error term
N total number of road segments

In Equation 2, the unobserved component y_n^* is associated with impacting factors, and based on y_n^*, the observed accident frequency y_n is associated with impacting factors as defined below:

$$y_n = \begin{cases} 0 & \text{if } y_n^* \leq 0\,(\text{low}) \\ 1 & \text{if } 0 < y_n^* \leq \mu\,(\text{medium}) \\ 2 & \text{if } y_n^* > \mu\,(\text{high}) \end{cases} \tag{3}$$

where, $\mu =$ positive threshold.

In Equation 3, the three coded responses 0, 1, and 2 represent the (un)safety categories discussed above.

The probabilities associated with the coded responses of the OPM are as follows:

$$P_n(0) = \Pr(y_n = 0) = \Phi(-x_n'\beta), \tag{4}$$

$$P_n(1) = \Pr(y_n = 1) = \Phi(\mu - x_n'\beta) - \Phi(-x_n'\beta), \tag{5}$$

$$P_n(2) = \Pr(y_n = 2) = 1 - \Phi(\mu - x_n'\beta). \tag{6}$$

where, $\Phi(.)$ standard normal cumulative distribution function.

Using these probabilities, the parameters β and μ can be estimated using the maximum likelihood method.

Taking into account the increasing nature of the ordered classes, the interpretation of the β parameters is as follows: positive signs indicate a higher risk as the value of the associated variable increases, whereas negative signs suggest the reverse. However, in probability models such as the OPM, the sign of β does not always determine the direction of the effect on the intermediate outcomes. In that sense, the marginal effects can provide a better interpretation of changes in the independent variables $\mathrm{X_n}$.

4. Estimation Results

The estimation results of the OPM and the marginal effects for each variable are presented in Table 2. The parameter estimates and the threshold parameter are significant at the 95% level, indicating a possible relationship between these variables and the occurrence probability of the three categories. The parameter estimates indicate the effect of the independent variables on the latent propensity of accident occurrence for the segments. The marginal effects represent the directionality and magnitude of the effects of the variables on the probabilities of each category. Note that the sum of the marginal effects is zero, which follows from the requirement that the probabilities add to one (Greene, 2008).

Table 2
Ordered probit model results

Parameter	Estimated Value	Standard Error	P[Z > z]	Marginal Effects Low	Medium	High
Constant	−9.984	0.505	0.0000	—	—	—
LnAADT	0.393	0.045	0.0000	−0.129	0.116	0.013
LnLength	1.107	0.047	0.0000	−0.364	0.328	0.036
TimeTrend	−0.056	0.022	0.0119	0.018	−0.017	−0.002
Minor intersections	0.038	0.007	0.0000	−0.012	0.011	0.001
Low density of buildings (LU2)	−0.287	0.082	0.0005	0.089	−0.081	−0.008
Industrial (LU3)	0.614	0.175	0.0005	−0.229	0.190	0.038
Community buildings (LU4)	−0.494	0.132	0.0002	0.138	−0.128	−0.010
Historic center (LU5)	0.299	0.093	0.0012	−0.104	0.092	0.013
Local distributor roads	−0.186	0.069	0.0076	0.061	−0.055	−0.006
μ	1.614	0.062	0.0000	—	—	—
Log-likelihood at zero	−1682.965					
Log-likelihood at convergence	−1158.103					
Percentage correctly predicted	73%					

LU = land use.

The marginal effects of the traffic volume (AADT), segment length, and minor intersection density are positive for Categories 1 (medium) and 2 (high), suggesting the likelihood that these variables are associated with a high risk of an accident occurring. These findings are in line with those reported by various studies based on count-data models (Greibe, 2003; Ivan et al., 2000; Karlaftis & Tarko, 1997; Mountain et al., 1996, 1998; Wedagama et al., 2006; Wier et al., 2009) as well as with the results obtained from a count-data model applied to the data set used in this study (Ferreira, 2010; Ferreira & Couto, 2011). The time trend effect represented by the time trend variable is negative for the two last categories (1 and 2), thus indicating an annual decline in accident frequency for the segments with higher risk.

The parameter estimates for LU3 (industrial area) and LU5 (historic center area) are positive and are thus associated with a higher risk of an accident occurring, with positive marginal effects for Categories 1 and 2. This is consistent with the findings of previous research (Greibe, 2003; Ivan et al., 2000) and is logical considering the local analysis. LU3 is only composed of one zone in the city although there are various trip attractors depending on the hour and the day of the week. During working hours, this zone is associated with the presence of heavy vehicles and goods deliveries for the industry sector there. On weekends and nonworking hours, this zone is related to risk behaviors associated with driving (alcohol, speeding, etc.) due to the fact that there are bars and restaurants in the area. This may explain the positive value of the parameter estimates, which demonstrates an increase in the probability of an accident occurring in the segments located in this type of land use area. Furthermore, in the case of LU5, the positive value of the parameter estimates may be related to the fact that this is a historic center area with a high-density building zone with old and outdated road infrastructures. In addition, there is also a high pedestrian volume and various forms of public transport. The negative values of the parameter estimates for LU2

Table 3
Cross tabulation of predicted versus actual observations

Actual Observations	Total Actual Observations	Predicted Values		
		0	1	2
0	1260	1101[a]	155	4
1	557	249	283[a]	25
2	163	12	99	52[a]
Total predicted values	1980	1362	537	81

[a]Observations correctly predicted.

and LU4 are also in line with expectations. In fact, zones with low densities of buildings, as is the case for LU2, are associated with a decrease in terms of accident risk. In addition, in Porto city, the community building areas (LU4) encompass two zones that include a hospital, a sports hall, and university buildings. The reduced accident risk in these zones may be explained by the low density of buildings with suitable road infrastructures and perhaps by highly seasonal movements associated with fewer traffic conflicts in this zone.

Moreover, the effect of road function classification on accident occurrence probability was also as expected. Thus, the negative value for the parameter estimates of the local distributor segments reveals a decrease in accident risk that may be associated with narrower streets, which promote better driving behavior (e.g., lower speeds).

Finally, the analysis of the distribution of the three category probabilities of the 396 segments used in this study indicates that, in 69% of the segments, the "low" category is more likely to occur (27% and 4% for "medium" and "high," respectively). For road network evaluation purposes, this kind of result can be compared to the results of other possible scenarios while maintaining the ability to identify segments that can be improved, based on a road safety perspective.

These results correspond to a correct prediction percentage of 73% (superscripted values in Table 3). Table 3 shows the accuracy of prediction for each category. The "high" category (represented in the table by the number *2*) has the highest rate of misclassification of observations predicted by the model, which may be due to the fact that there are fewer observations for this category, thus resulting in heterogeneity phenomena.

4.1. Analysis of Hypothetical Scenarios

Transportation planning is a cyclical process that continually improves the current transportation system through planning and programming project and nonproject solutions to the needs of the system. A project solution is a physical improvement to the system infrastructure; a nonproject solution is a management operation such as a traffic management. Transportation planners examine demographic characteristics and travel patterns for a given area and then analyze alternatives for the area's transportation system. This is provided to the decision makers who evaluate the alternatives to determine the most suitable system that meets the future needs. Therefore, the analysis process should evaluate and optimize land use and transportation network proposed to achieve not only mobility and/or air quality, but safety as well (AECOM et al., 2002). Land use planning can be described as the planning of the relative location of different types of land use and of the way they are connected

(Hummel, 2001). Different types of land use generate different numbers of trips and engender different driver behaviors, which in turn have the potential to cause road accidents (Berkovitz, 2001; Wedagama et al., 2006). Therefore, if the safety effects of decisions in land use planning are considered, planners are informed of the consequences, and thus developments can be directed toward a safer direction. Similarly, it is recognized that the general risk for traffic accidents varies by highway functional class. Actually, highway functional class is usually connected to the spatial organization of land use types, being for that common to be planned altogether.

This section presents a hypothetical planning exercise, which intends to replicate commonly arisen situations under the planning process as described above, to demonstrate the application of the OPM. Therefore, two sample cases were set using as reference scenarios (S_0) the attributes of actual segments selected from the Porto's road network database. Then, hypothetical scenarios were created considering commonly planning situations at a short term and a long term. In the first sample case, a road function reclassification was set as a short term planning situation, by changing the reference scenario (S_0) of the segment ID1, whose characteristics are described in Table 4. The planned future situation for segment ID1 (S_1) comprehends a physical change by enlarging the sidewalks, which involves changes in the traffic flow as a result of an one-way traffic elimination (two-ways for one-way), estimated in a decrease of the traffic volume of around 40%.

The second proposed sample case represents a long term planning exercise (10 years longer than the reference year), considering the case of the segment ID2 located in an area with low density of buildings (Table 4). The hypothetical future scenario (S_1) forecasts a growth in residential building construction which changes the initial land use classification LU2 to LU1, along with an increase of the traffic volume of around 100%. Table 4 shows the attributes of the segments at the reference scenario (S_0) and at the planned scenario (S_1) for both sample cases. The OPM was applied to these sample cases, whose results are shown in Table 4. As can be seen, the accident risk of the segment ID1 has decreased due to the planned actions (S_1), because the probability associated to the high degree of hazardous (category 3) has decreased. Despite the accident risk has decreased, the S_1 scenario still has a high accident risk, and therefore, planners should pay special attention to this future road intervention, which should be accompanied by safety measures to improve safety. In addition, Table 4 shows that the planned scenario set to the segment ID2 (S_1) produces a slight change in the probability of the "medium" degree of hazardous when compared with the reference scenario (S_0), although keeping it as the most likely. Despite the increase in the building construction followed by an increase in the traffic volume, the accident risk of the simulated scenario (S_1) has not changed, because, in a long-term forecast, a decrease of accident frequency resulting from the time trend effect is expected to occur.

Additionally, actual accident figures and the number of accidents estimated by a count-data model (NB model) applied to the same segment scenarios are presented in Table 4. Note that these latter values intend to demonstrate that, despite the fit performance of the models, the estimated number of accidents is less suitable for a wide level of analysis such as the transportation planning process.

5. Summary and Conclusions

A common technique used for safety planning models is the GLM procedure with the assumption of a NB or Poisson error distribution. With this technique, the dependent

Table 4
Two examples of hypothetical scenarios

Sample Cases	Actual Number of Accidents (5-year Average)	Prob (y = 0)%	Prob (y = 1)%	Prob (y = 2)%	Y[a] (Accident Number)
Segment ID1:					
S_0: Reference segment scenario-AADT = 21673; Length = 351; 8.6 minor inter. per km; LU1; principal distributor road	7.4	2	33	65	5.6
S_1: −40% of AADT (13004); local distributor road	—	9	51	40	3.9
Segment ID2:					
S_0: Reference segment scenario-AADT = 12643; Length = 352; 8.5 minor inter. per km; LU2; principal distributor road	6.2	11	54	35	3.3
S_1 (10 years long): +100% of AADT (25285); LU1	—	10	53	37	3.5

AADT = annual average daily traffic.
[a]These predicted values were based on a NB model, of which, the parameter estimates were statistically significant with $R^2_{FT} = 67\%$ (Ferreira, 2010).

variable is usually the number of accidents per TAZ or per segment and intersection in the case of area-level or road-level models, respectively. Thus, the model outcome is a quantitative response. However, at the road-planning level, there is a lack of data for properly assessing safety by predicting the number of accidents because such a value is associated with a series of factors that are unknown at this point. Furthermore, the main point of a safety planning model is to compare and evaluate alternative solutions rather than to predict/forecast a number.

In this sense, this work presents an alternative approach based on a qualitative response model. Three responses were defined to reflect different categories based on a range of accident numbers that can be associated with degrees of hazard. The methodology for defining the response is flexible and can be adjusted to other jurisdictions. Thus, an OPM was applied to estimate the parameters and compute the marginal effects. All of the parameter estimates were statistically significant, and the marginal effect values were in line with findings reported by several count-data models and with the results obtained from a count-data model applied to the same data set used in the present study. Moreover, the results of the case studied demonstrate that, in 69% of the segments, the "low" category, defining a low degree of hazard, is more likely to occur (27% and 4% for "medium" and "high," respectively), with a correct prediction percentage of 73%. In addition, an analysis of hypothetical scenarios for two sample cases was presented to illustrate an application of the OPM in typifying an accident risk analysis of possible planned scenarios. Hence, the probability of each response (category) occurring was computed, taking into account the attributes of two actual segments selected from Porto's road network and the correspondent planned attribute changes (simulated scenarios). The objective of this planning exercise is to simulate hypothetical applications to exemplify the use of the OPM as a safety tool for planners. Moreover, the effect of changing a road attribute can be related with the three degrees of hazardous.

Based on these issues, it can be concluded that the OPM presents an alternative approach as a safety tool for urban planners, providing a risk analysis with a simple and realistic interpretation of the variable effects. The OPM outcomes can be used among other traditional evaluation criteria in the strategic planning process, thus enabling safety to be included. Note that the concept of this approach can be applied to an aggregate model. It should be stressed that the methodology presented in this study is especially appealing to the application for a priori analysis, when there is not actual accident data.

This research represents a step toward an appropriate and reliable safety analysis at the planning level. However, further research should be done, focusing on applying alternative discrete choice models to compare the ordered response mechanism and the unordered response mechanism (represented by, e.g., a multinomial logit model) to select the more appropriate mechanism.

Notes

1. For other purposes, such as hot-spot identification, it may be useful to use risk factors as the categorical dependent variable.
2. Other options for category definitions were analyzed, including categories that differentiate accident severity (0 accidents, injury accidents, no-injury accidents). However, based on the statistical significance of the parameter estimates, these options were rejected.
3. Segment length was defined as part of a road network data set used under a doctoral study (Ferreira, 2010) where the influence area of an intersection was 15 meters from the center of the intersection.

4. Correlations among the variables were analyzed using a correlation matrix. This allows one to assume that the explanatory variables are not correlated ($\rho \leq 0.3$) (Ferreira, 2010).

References

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research 34*, 597–603.

AECOM, Bellomo-McGee Inc. and Ned Levine & Associates. (2002). *Considering safety in the transportation planning process*. Washington, D.C.: Federal Highway Administration.

Autoridade Nacional de Segurança Rodoviária (ANSR). (2009). *Sinistralidade rodoviária. Elementos estatísticos de 2009* [Road Safety. Statistical Data-2009]. Lisbon, Portugal: Observatório de Segurança Rodoviária, ANSR.

Berkovitz, A. (2001). The marriage of safety and land-use planning: A fresh look at local roadways. *Public Roads*, *65*(2). Retrieved from http://www.fhwa.dot.gov/publications/publicroads/01septoct/marriage.cfm

Bhat, C. R., & Pulugurta, V. (1997). A comparison of two alternative behavioral choice mechanism for household auto ownership decisions. *Transportation Research Part B 32*(1), 61–75.

Carson, J., & Mannering, F. (2001). The effect of ice warning signs on ice-accident frequencies and severities. *Accident Analysis and Prevention*, *33*, 99–109.

Commision of the European Communities. (2009). *Action plan on urban mobility (Draft)*. Brussels, Belgium: Commission of the European Communities.

Dissanayake, D., Aryaija, J., & Wedagama, D. M. P. (2009). Modelling the effects of land use and temporal factors on child pedestrian casualties. *Accident Analysis and Prevention*, *41*, 1016–1024.

Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered responde model for examining pedestrian and byciclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, *40*, 1033–1054.

Ferreira, S. (2010), *A Segurança Rodoviária no processo de planeamento de redes de transportes* [Road safety in the planning process of road networks] (Unpublished doctoral dissertation), University of Porto, Porto, Portugal.

Ferreira, S., & Couto, A. (2011). Urban road planning: a safety perspective. In *Proceedings of the 1st International Conference on Transportation Information and Safety*, Vol. 1: Highway Transportation. ASCE Proceedings (pp. 622–628). Wuhan, China:

Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Thomsen, L. K. (1995). Measuring the contribution of randomness, exposure, weather and dayligth to the variation in road accident counts. *Accident Analysis and Prevention*, *27*(1), 1–20.

Greene, W. H. (2008). *Econometric analysis*. NJ: Pearson International.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention*, *35*, 273–285.

Guevara, F. L. d., Washington, S. P., & Oh, J. (2004). Forecasting crashes at the planning level. *Transportation Research Record*, *1897*, 191–199.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2003). Macro-level accident prediction models for evaluating the safety of urban transportation systems. *Transportation Research Record*, *1840*, 87–95.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2010). Development of planning level transportation safety tools using geographically weighted poisson regression. *Accident Analysis and Prevention*, *42*, 676–688.

Hummel, T. (2001). *Land use planning in safer transportation network planning*. Leidschendam, The Netherlands: Institute for Road Safety Research.

Ivan, J. N., Wang, C., & Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention*, *32*, 787–795.

Karlaftis, M. G., & Tarko, A. P. (1997). Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention*, *30*(4), 425–433.

Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: An application of ordered probit models. *Accident Analysis and Prevention*, *34*, 313–321.

Leur, P. D., & Sayed, T. (2002). Development of a road safety risk index. *Transportation Research Record*, *1784*, 33–42.

Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: II. *Zonal generators. Accident Analysis and Prevention*, *27*(5), 675–685.

Lord, D., & Persaud, B. N. (2004). Estimating the safety performance of urban road transportation networks. *Accident Analysis and Prevention*, *36*, 609–620.

Mountain, L., Fawaz, B., & Jarrett, D. (1996). Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, *28*(6), 695–707.

Mountain, L., Maher, M., & Fawaz, B. (1998). The influence of trend on estimates of accidents at junctions. *Accident Analysis and Prevention*, *30*(5), 641–649.

Naderan, A., & Shahi, J. (2010). Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis and Prevention*, *42*, 339–346.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, *12*(2), 105–113.

Qi, Y., Smith, B. L., & Guo, J. (2007). Freeway accident likelihood prediction using a panel data analysis approach. *Journal of Transportation Engineering*, *133*(3), 149–156.

Savolainen, P. T., Mannering, F., Lord, D., & Quddus, M. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis and Prevention*, *43*(5), 1666–1676.

Sayed, T., & Lovegrove, G. R. (2006). Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Canadian Journal of Civil Engineering 33*(5), 609–621.

Tarko, A. P., Inerowicz, M., Ramos, J., & Li, W. (2008). Tool with road-level crash prediction for transportation safety planning. *Transportation Research Record*, *2083*, 16–25.

Tavares, J. P. (2003), *Aplicabilidade e robustez de modelos de afectação de tráfego em redes urbanas* [Applicability and robustness of traffic assignment/simulation models in urban roads] (Unpublished doctoral dissertation), University of Porto, Porto, Portugal.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press.

Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis and Prevention*, *40*, 1674–1682.

Washington, S., Schalkwyk, I. V., Meyer, M., Dumbaugh, E., & Zoll, M. (2006). Incorporating safety into long-range transportation planning. NCHRP Report 546, Washington, D.C.: Transportation Research Board. pp. 158.

Wedagama, D., Bird, R., & Metcalfe, A. (2006). The influence of urban land-use on non-motorised transport casualties. *Accident Analysis and Prevention*, *38*, 1049–1057.

Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis and Prevention*, *41*(1), 137–145.