Accident prediction models in urban areas: Lisbon case study

S. Vieira Gomes[1], C. Carvalheira[2], J. Cardoso[1] & L. Picado Santos[2]
[1]LNEC – Laboratório Nacional de Engenharia Civil, Portugal
[2]Department of Civil Engineering, University of Coimbra, Portugal

Abstract

According to official statistics, an important percentage of accidents and injuries are reported in Portuguese urban areas: from 2004 to 2007, 70% of the injury accidents and 43% of the fatalities occurred inside urban areas. To develop an efficient and affective strategy towards road safety, it is necessary for road administrations to have the proper tools for the quantification of safety levels and the explicit consideration of safety issues in the road management process. The analysis of spatial accident distribution in road networks and the knowledge of the relations between accident frequencies and variables describing the urban road environment will allow a more efficient definition of priorities for intervention and safety funding. This can be achieved by means of accident prediction models adapted to the urban context where they are applied, and by the use of a Geographic Information System based methodology to analyse spatial patterns of road accidents. This paper summarizes the result of the bibliographic study on accident prediction models applied to urban areas and the data collection on road accidents, road infrastructure characteristics and traffic and land use information, integrated in a Geographic Information System, where it may be graphically visualized and analyzed.

Keywords: accident prediction model, road safety, pedestrians, geographical information system.

1 Introduction

The estimation of the number of accidents that may occur on a given road section as a result of its design characteristics plays an important role in the highway engineering community. Several studies were developed in this area,

aiming to determine the effects of different design elements on road safety. Accident prediction models allow for a realistic estimate of the expected number of accidents or victims, as a function of explanatory variables such as the traffic volumes and road geometry characteristics. They represent relations between independent variables and accidents, through mathematical functions, allowing one to quantify the variation in the level of safety associated with changes in each considered variable. The development of these estimates is a fundamental component for safety considerations in road planning, since it allows the understanding of accident factors, the selection of suitable safety countermeasures based on these factors and the evaluation of the effectiveness of the applied engineering measures.

2 Accident prediction models

2.1 Basis for development

The development of accident prediction models is a meticulous task, so that the results and interpretations that they provide are both realistic, useful and the information they require is parsimonious.

Several recommendations for the development of accident prediction models were recently proposed in the RIPCORD-ISEREST project [1] of the 6th framework program:

1. The probabilistic distribution of accidents in the original data set must be identical to the one of the residual terms of the model.
2. Models must be disaggregated by level of severity (fatal accidents, accidents with victims and property-damage-only accidents), by type of road element (road section, intersections, bridges, tunnels, curves and railroad crossings) and by class of vehicles (trucks, cars, two wheelers, pedestrians and cyclists).
3. The correlations between the explanatory variables must be analyzed in detail, with justification for the functional forms chosen, as well as all the causal relations. All the variables with high correlation between them, as well as the ones that are considered confounding, must be eliminated. The possibility of omitted variable bias must be taken into account, since it is not feasible to create an accident prediction model with all the variables that influence accident occurrence.
4. The overall goodness-of-fit of the model must allow the decomposition of the variation of the number of accidents in: b) random variation, b) systematic variation explained by model, and c) systematic variation not explained by the model. This last one must be analyzed, to decide if the over-dispersion can be described by a simple parameter or if it must be modelled by a variable parameter.
5. The predictive performance of the model must be tested through its application on a data set that has not been used for its development.

2.2 Types of models

Accident prediction models can be classified in different ways [2]:

- According to the technique used to estimate the effect on safety indicators:
  - **Before-after studies**, that include on its methodology mathematical expressions that allow estimating the effect on safety of changes in the transport system. They are considered quite efficient as long as the disturbing factors are controlled and the sample dimension and the analysis techniques used are adequate. However, the applicability of the defined relations is restricted, since they are specific for the context where they were adjusted, a fact that imperils its generalization.
  - The **adjustment of mathematical equations** through statistical methods, allows relating data on accidents or victims with a series of explanatory variables, creating the so called accident prediction models. The main advantage of this type of models relies on the possibility for a direct use in the evaluation of the effect on the safety indicator of changes on the exploratory variables. Its usefulness increases when the number of explanatory variables is high, when the number of confounding variables is high (and they cannot be treated through the consideration of control groups) or when the sample (of accidents or victims) is small.
- According to the consideration of "time" in the model: parametric (cross-sectional models) or variable (time series models):
  - **Cross-sectional models** allow representing the variation between variables that characterize different road entities and its level of safety, for the same period. These models explore the variation between different entities in the same time period, relating accidents to the variation of characteristics of different entities (any geographic unit or physical element – people, vehicles or groups with similar characteristics). In the development of this type of models it is important to ensure that the road entities are similar and that all variables with influence on accident occurrence are considered. This type of models is normally applied when data sets of considerable dimension are available and when the possible explanatory variables are independent and with low co-variation.
  - **Time series models** comprise several observations of the same element in time. In this type of modelling, variations between consecutive observations are very small, especially when there is a considerable colinearity between potential explanatory variables. In these models, the inclusion of all the relevant explanatory variables is generally more difficult. For the development of this type of models several tools are available to deal with problems of self-correlation and auto-regression, among others [2–4].

- According to level of disaggregation of the variables used: aggregated and disaggregated models:
  - **Aggregate models** allow the description of general safety trends on the regional or national level, making possible the development of short-term safety estimations, as a function of traffic and macro-economic variables. Estimates can be improved by including descriptive factors of the impact of safety measures. This type of models does not allow, however, to evaluate the effect of changes in parts of the transportation system, neither the global impact of safety interventions on specific groups of users, since they contain only broad macro-economic variables.
  - **Disaggregated models** may be used to represent the effect of changes in specific parts of the transportation system or in the safety of specific user groups. They are used in the evaluation of safety policies and in road safety estimations, being therefore considered an important instrument to road safety management at a macroscopic level (regional or national). The usual types of disaggregation are by transport mode, age group, sex and type of road [2, 3].

2.3 Accident prediction models in intersections and road links

This section summarizes the result of the bibliographic study on accident prediction models applied to urban areas. Commonly, different models are developed for intersections and road links. Several types of models where analyzed, namely for accidents involving pedestrians or cyclists, collisions between motorized vehicles, total accidents, non injury accidents, accidents with fatalities, injury accidents, night time accidents and accidents involving only vehicles. Models with other types of disaggregation were also studied.

2.3.1 Aggregated models - regional level

Washington et al. [5] developed accident prediction models for: total accidents, fatal accidents, fatal and injury accidents, injury accidents, pedestrian accidents, cyclist's accidents, night time accidents and accidents without victims. The standard form off all models is a log linear regression model, which included general variables regarding: population (total, by age groups, by area, by means of transportation, etc), road length (total, main roads, motorways, urban/rural roads, etc), vehicles kilometres travelled, intersections density, average income and number of housing units (total and per area):

$$Log(Accidents+1) = a_0 + a_1 \times Variable_1 + a_2 \times Variable_2 + ...$$

2.3.2 Disaggregated models

Several authors have developed accident prediction models specifically for pedestrian, since these type accidents are a major concern in urban areas. Turner et al. [6] developed several accident prediction models regarding specifically pedestrians in intersections (signalized intersections, roundabouts and T-junctions):

$$Accidents_{pedestrians} = a_0 \times Traffic_1^{a_1} \times Traffic_2^{a_2} \times Pedestrians^{a_3}$$

They disaggregated their multiplicative models by type of movement: crossing, left-turn, and right-turn, using traditional variables (motorized traffic volumes and pedestrian volumes), and used specific variables associated with conflicting movements, namely: the proportion of pedestrians that cross with the "green-man", the average crossing distance and number of lanes that vehicles that turn left have to cross:

$$\text{Accident}_{\text{pedestrians}} = a_0 \times \text{Traffic}^{a_1} \times \text{Pedestrians}^{a_2} \times \text{Crossing}_{\text{dist}}^{b_3} \text{ (for example)}$$

Pedestrian accidents at intersections were also modelled by other authors, namely Brüde and Larson [7], Maher and Summersgill [8] and Gårder [9] (for roundabouts). All of them used multiplicative models that included motorized vehicles and pedestrian volumes as explanatory variables:

$$\text{Accident}_{\text{pedestrians}} = a_0 \times \text{Traffic}^{a_1} \times \text{Pedestrians}^{a_2} \text{ [7, 8]}$$

$$\text{Accident}_{\text{pedestrians}} = a_0 \times (\text{Traffic} \times \text{Pedestrians})^{a_1} \text{ [9]}$$

Maher and Summersgill developed accident prediction models for pedestrians in urban T-junctions without median, with a disagregation of motorized traffic by major and minor leg:

$$\text{Accident}_{\text{pedestrians}} = a_0 \times \text{Traffic}_1^{a_1} \times \text{Traffic}_2^{a_2} \times \text{Pedestrians}^{a_2}$$

Additionally they developed accident prediction models for other types of accidents in the same type of intersections, namely: total accidents, property damage only accidents. They also developed accident prediction models for road links, namely: total accidents, pedestrian accidents and property damage only accidents. The general form of the mentioned models is:

$$\text{Accident}_{\text{pedestrians-road links}} = a_0 \times \text{Road Lenght} \times \text{Traffic}^{a_1} \times \text{Pedestrians}^{a_2}$$

Total accident frequency is the most common response variable in accident modelling. Several authors developed mathematical functions to explain total accident occurrence at intersections. The common explanatory variables used were also motorized vehicles and pedestrian volumes, sometimes disaggregated by major and minor legs:

- Leden [10] (intersections)

$$\text{Total accidents} = a_0 \times \text{Traffic}^{a_1} \times \text{Pedestrians}^{a_2}$$

- Sayed and Rodriguez [11] (non-signalized urban intersections controlled by STOP signs), Greibe [12] (urban intersections with three or four legs with and without traffic signals) and Mountain and Fawaz [13] (intersections with different types of traffic control)

$$\text{Total accidents} = a_0 \times \text{Traffic}_1^{a_1} \times \text{Traffic}_2^{a_2}$$

(Only injury accidents for [11] and [13])

- Lord and Persaud [14] (signalized urban intersections with four legs) and Persaud et al. [15] (three or four legs intersections with and without signals)

$$\text{Total accidents} = a_0 \times \text{Traffic}_1^{a_1} \times \text{Traffic}_2^{a_2} \times e^{a_3 \times \text{Traffic}_2}$$

- Bauer and Harwood [16] (collisions in urban and rural intersections - four legs with STOP; three legs with STOP and four legs with signal lights)

$$\text{Total collisions} = e^{a_0} \times \text{Traffic}_1^{a_1} \times \text{Traffic}_2^{a_2} \times e^{a_3 \times \text{Traffic}_2 + ... + a_n \times \text{Traffic}_n}$$

Accident prediction models in road links were also the scope of several authors, mainly in what concerns total accidents. Motorized vehicle traffic volumes were the basic explanatory variables used, but other variables like road length, driveway density, number of minor intersections, pedestrian traffic volumes, road width, number of lanes and traffic speed were also used:

- Turner et al. [17]

$$\text{Total accidents} = a_0 \times \text{Traffic}^{a_1}$$

- Mountain et al. [18]

$$\text{Total accidents} = a_0 \times \text{Traffic}^{a_1} \times \text{Road length}^{a_2}$$

- Bonneson and McCoy [19]

$$\text{Total accidents} = \text{Traffic}^{a_0 + a_1 \times \text{Variable}_1 \times \text{Variable}_2} \times \text{Road length}^{a_2} \times e^{a_3 + a_4 \times \text{Variable}_3 + \ldots + a_n \times \text{Variable}_n}$$

- Abo-Qudais [20]

$$\text{Total accidents} = a_0 \times \text{Variable}_i^3 \times \text{Variable}_i^2 \times \text{Variable}_i$$

$$\text{Total accidents} = a_0 \times \text{Speed}^2 \times \text{Speed}$$

$$\text{Total accidents} = a_0 \times \text{Traffic}^{a_1}$$

- Greibe [12]

$$\text{Total accidents} = a_0 \times \text{Traffic}^{a_1} \times a_2 \times \ldots \times a_n$$

3 Geographical Information System (GIS) accident database

The present study is part of the "IRUMS – Safer Roads in Urban Areas" project, carried out at the National Laboratory of Civil Engineering and at the Department of Engineering of the University of Coimbra, financed by the Foundation for Science and Technology. This project intends to develop methods for safety management of urban road networks. Procedures for the estimation of expected accident frequencies, identification of sites with a promise and selection of efficient corrective measures are being developed as well. The case study is being applied in Lisbon.

Portuguese accident's register is made by the police, which is called for every injury accident, but in case of property damage only accidents, where it is up to the accident's participants to decide if they what the presence of the police or not. This leads to a percentage of under reported accidents not quantified in Portugal.

Major characteristics of each accident are reported by police officers on a special form. Afterwards, a statistic form is filled, based on the previous one, and sent to the National Road Safety Authority, which adds the relevant information to the national accident database. Although it may be considered a substantial database, for urban accident analysis it is insufficient, as detailed data on the accident location is lacking, as only the name of the road is included. To overcome this problem, it was necessary to lock up in the first accident bulletins (filled by the police by the time of the accident), which included the accident description and sketch, in order to find a spatial reference that could help on the

accident spatial location. Although this is only being made for the city of Lisbon, several years were collected (2003-2007), leading to a substantial data collection.

In order to achieve preliminary conclusions on spatial accident distribution, the geocoding process was made backwards (2007 to 2003). The next picture shows the ongoing work with some of the 2007 accidents already located.
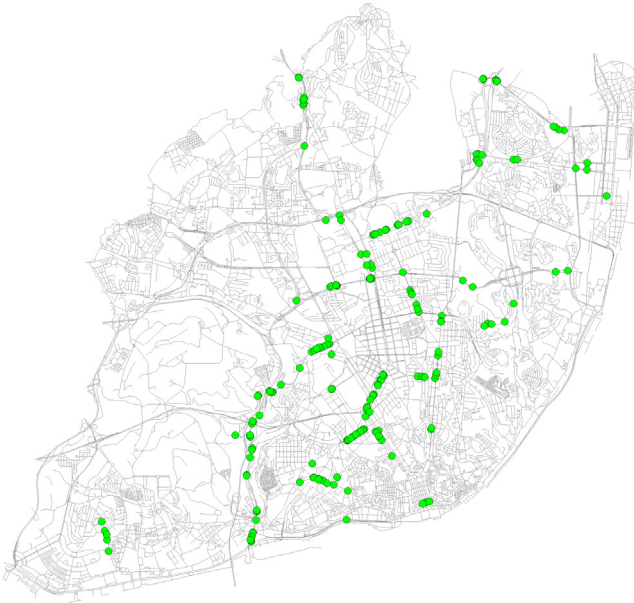

Figure 1:   Lisbon accidents distribution in 2007 (partial).

The Lisbon accident prediction model final target also implies traffic data collection. Several locations were already defined for traffic counts, based on this first accident distribution.

Identification and ranking of high accident frequency locations will be made by studying the spatial patterns of expected accident frequency distributions. This can be achieved through the use of several methods to create concentration maps, based on density values, for instance the Simple Method and the Kernel Method [21]. The GIS based methodology has the advantage of reducing the subjectivity in the analysis process, allowing decision makers to use the obtained results for adequate safety fund allocation.

4   Final notes

This paper summarizes the results of the bibliographic study on accident prediction models applied to intersections and road links in urban areas. Accident prediction models are mathematical functions that describe the relation between the road safety and explanatory variables, such as traffic, road length

and carriageway width, number of intersections, etc. Its most common form is expressed as the following multiplicative expression: $A = \alpha \times T_1^\beta \times T_2^\beta \times e^{\sum \gamma_i \cdot x_i}$, where A is the expected number of accidents, which is a function of traffic volumes (T_i) and with other factors (X_i). The effect of traffic in the accident's occurrence is modelled through the power ß. The effect of the several risk factors that usually influence accident frequency is modelled through an exponential function of base e and raised to the sum of the product of the γ_i coefficients by the risk factors, x_i.

In disaggregated accident prediction models for intersections, the most common explanatory variables used are traffic volume, and for accidents involving pedestrians, the average distance crossed or lane with, which reveals its high significance in the explanation of these phenomena. Regarding road links, traffic volumes, road width, speed and road length were the most frequent significant explanatory variables used to model road accidents.

Considerable progress has been made in the techniques for establishing the relationship between accidents, traffic volumes and road geometry. Specific problems such as low mean value, overdispersion, disaggregation of data over time and random errors were already identified by several authors and the solutions that they outlined will help in the development of accident prediction models for Portuguese urban areas, adjusted to data from the city of Lisbon.

The GIS Accident database in construction for the city of Lisbon (2003-2007) will allow the study of the spatial patterns of accidents distributions, and identification and ranking of high expected accident frequency, with tools that reduce the subjectivity in the analysis process, allowing to contribute to an effective and efficient strategy of road safety enhancement.

References

[1] Reurings, M.; Janssen, T.; Eenink, R.; Elvik, R.; Cardoso, J.; Stefan, C. - Accident Prediction Models and Road safety Impact Assessment: a state-of-the-art. European project RIPCORD-ISEREST, June of 2005.

[2] Cardoso, J. L. – Métodos racionais de apoio à intervenção da engenharia em segurança rodoviária. LNEC, Lisboa, 2007.

[3] European Commission (EC). 2004. COST Action 329: Models for Traffic and Safety Development and Interventions. Luxembourg: European Communities.

[4] OECD - Road safety principles and models: Review of descriptive, predictive, risk and accident consequence models, 1997.

[5] Washington, S.; Meyer, M.; Schalkwyk, I.; Dumbaugh, E.; Mitra, S.; Zoll, M. – Guidance: Incorporating safety into long-range transportation planning. National Cooperative Highway Research Program Report 546, Transportation Research Board of the National Academies. Washington, D. C., 2006.

[6] Turner, S. A.; Roozenburg, A. P.; Francis, T. – Predicting Accident Rates for Cyclists and Pedestrians – Land Transport New Zealand Research Report 289, Christchurch, New Zealand, 2006.

[7] Brüde, U.; Larson, J. – Models for predicting accidents at junctions where pedestrians and cyclist are involved. How well do they fit? – Accident Analysis and Prevention, Vol. 25, No. 5, p. 499-509, 1993.

[8] Maher, M. J.; Summersgill, I. – A comprehensive methodology for the fitting of predictive accident models – Accident Analysis and Prevention, Vol. 28, no. 3, p. 281-296, 1996.

[9] Gårder, P.E – The impact of speed and other variables on pedestrian safety in Maine – Accident Analysis and Prevention, Vol. 36, p. 533-542, 2004.

[10] Leden, L. – Pedestrian Risk decreases with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario – Accident Analysis and Prevention, Vol. 34, p. 457-464, 2002.

[11] Sayed, T.; Rodriguez, F. – Accident prediction models for urban unsignalized intersections in British Columbia – In: Transportation Research Record 1665, p. 93-99, 1999.

[12] Greibe, P. – Accident prediction models for urban roads. In: Accident analysis and prevention, 35, p. 273-285, 2003.

[13] Mountain, L.; Fawaz, B. – Estimating accidents at junctions using routinely-available input data – Traffic Engineering &Control, 37 (11), p. 624-628, 1996.

[14] Lord, D.; Persaud, B.N. – Accident prediction models with and without trend. Application of the generalized estimating equations procedure – In: Transportation Research Record 1717, p. 102-108, 2000.

[15] Persaud, B.; Lord, D.; Palmisano, J. – Calibration and transferability of accident prediction models for urban intersections – In: Transportation Research Record 1784, p.57-64, 2002.

[16] Bauer, K. M.; Harwood, D. W. – Statistical models for at-grade intersection accidents – Addendum. FHWA-RD-99-094 Report, U.S. Department of Transportation, Federal Highway Administration, 2000.

[17] Turner, S.; Durdin, P.; Bone, I.; Jackett, M. – New Zealand accident prediction models and their applications. In: Transport: our highway to a sustainable future: proceedings of the 21st ARRB and 11th REAAA Conference, Cairns, Queensland, Australia, 18-23 of May 2003.

[18] Mountain, L.; Fawaz, B.; Jarret, D. – Accident prediction models for roads with minor junctions. In: Accident Analysis & Prevention, 28 (6), p. 695-707.

[19] Bonneson, J. A.; McCoy, P. T. – Effect of median treatment on urban arterial safety; an accident prediction model. In: Transportation Research Record nº 1581, p. 27-36, 1997.

[20] Abo Qudais S., "Urban road accident prediction models", Roads-Routes, AIPCR, World Road Association, n. 309-I, 2001.

[21] Pulugurtha, S. S., Krishnakumar, V. K., Nambisan, S. S. – New methods to identify and rank high pedestrian crash zones: An illustration. In: Accident Analysis & Prevention, 39 (2007), p. 800-811.