THE VARIABILITY OF URBAN SAFETY PERFORMANCE FUNCTIONS FOR DIFFERENT ROAD ELEMENTS: AN ITALIAN CASE STUDY

Paolo Intini
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Polytechnic University of Bari, Bari, Italy, 70125
Email: paolo.intini@poliba.it

Nicola Berloco
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Polytechnic University of Bari, Bari, Italy, 70125
Email: nicola.berloco@poliba.it

Gabriele Cavalluzzi
Master's Thesist
Polytechnic University of Bari, Bari, Italy, 70125
Email: cavalluzzigabriele@gmail.com

Pasquale Colonna
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Polytechnic University of Bari, Bari, Italy, 70125
Email: pasquale.colonna@poliba.it

Dominique Lord
Zachry Department of Civil and Environmental Engineering
Texas A&M University, College Station, TX, United States, 77843
Email: d-lord@tamu.edu

Vittorio Ranieri
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Polytechnic University of Bari, Bari, Italy, 70125
Email: vittorio.ranieri@poliba.it

Word Count: 6,745 words (including abstract, manuscript, acknowledgements, references) + 3 tables = 7,495 words

Submitted [August 1st, 2019]

ABSTRACT

Safety performance functions are used to predict crash frequencies based on several possible variables which include at least traffic volumes, geometric, and traffic-control variables. In urban environments, safety predictions are usually differentiated for homogeneous road elements: segments and intersections. Further disaggregations are often considered, such as one-way/two-way, one-lane/multilane segments, three/four-legged, signalized/unsignalized intersections.

In the context of a National research project, data about crashes, traffic, geometric, traffic-control and additional variables were collected for the road network of the City of Bari, Italy. 320 homogeneous segments and 120 intersections were included in the sample of sites, on which more than 1,500 fatal+injury crashes have occurred in a 5-years period (2012-2016). The study was conceived for research purposes and for being useful for practitioners.

The main research questions concerned: a) finding the best possible subsets for segments and intersections for safety modelling purposes, by discussing the related problems, and inquiring into the variability of predictors within subsets; b) comparing the modelling results with existing literature to highlight common trends and/or main differences; c) assessing the importance of additional crash predictors, besides traditional variables.

As a result of the study, six detailed models were developed for: one-way/two-way homogeneous segments, three/four-legged, signalized/unsignalized intersections. Crash predictors greatly vary within the different subsets considered. The effect of vertical signs on minor roads/driveways, critical sight distance, cycle crossings, pavement/markings maintenance was specifically discussed. Some common trends but also notable differences in both types and effect of crash predictors were found by comparing results with relevant literature.

Keywords: Safety Performance Functions, Crash predictions, Urban segments, Urban intersections.

INTRODUCTION

The use of Safety Performance Functions (SPFs) is crucial for road safety purposes, for making choices concerning different project alternatives on new or existing highways. Several functions were developed for different types of road elements, either road segments or intersections (e.g., *1-6*), in rural and urban environments. Most of these studies were conducted in North America, with a few applications in Europe, especially for urban areas. While functions developed in other contexts may be calibrated for being applied in other countries/regions (*6, 7*), the transferability of models is not without issues (*8, 9*).

Different aspects related to driving behavior, cultural and geographic variables (*6*) may affect the transferability of models in other contexts as they are. The use of a single calibration factor for taking into account these differences may not solve all transferability issues (6, *10*). In fact, some variables (e.g., traffic volumes, geometric characteristics) may influence crash occurrence in different ways according to diverse geographic contexts and road networks. Given this, it follows that the development of local SPFs may be of great importance, especially in countries where their practical use is still not largely considered, such as in Italy (*11*).

Previous modelling efforts have led to some European urban predictive models, e.g., for urban arterial segments and intersections in Denmark (*1*); urban intersections in Portugal (*2*); urban segments and intersections in Italy (*12, 13*). Some other studies were focused on specific crashes, such as to vulnerable users (e.g., *14*). However, some of these studies are old, limited to specific road elements (e.g., roundabouts, segments or intersections), or they use a limited number of predictors of road crashes.

A National research project (Pa.S.S.S. "Scientific Park for Road Safety") funded by the Italian Ministry of Infrastructures and Transport and focused on urban road safety and sustainable mobility is currently led by the City of Bari (Italy), main scientific partner: Polytechnic University of Bari. Among the project activities, the development of local SPFs was conducted.

Research questions

This study reports the modelling efforts undertaken for developing urban SPFs for different road elements. Besides being practically useful for road safety practitionners, the study was conceived for research purposes. In particular:

- The modelling effort was conducted for different road elements in the same area, to provide an organic portrait of crash performances of both segments and intersections, with the related influential variables. Morever, segments and intersections were further divided into sub-elements with homogeneous characteristics. The best subsets for modelling purposes were searched, by dividing segment or intersection road elements into couples of homogeneous subsets: two-way/one-way segments, one-lane/multilane segments, signalized/unsignalized intersections, three-legged/four-legged intersections.
- The comparison of the significant predictors resulted from the modelling stages with the significant crash predictors retrieved in previous research can reveal specific local differences which may be of interest for further studies.
- The influence of several road geometric and traffic control variables on road crashes was explored. Besides the typical variables used in similar studies, some other variables based

on visual ispections and geometric reconstructions of the sites inquired were taken into account, which are often not included in predictive models.

Details about the methods employed for this experimental study (data used, variables considered, analysis techniques) are provided next. Results are then presented and discussed. Finally, some conclusions are drawn, based on the research questions posed in this section.

METHODS

The experimental study aims at developing urban SPFs starting from datasets of crashes, traffic volumes and other road-related variables. The data collected, the variables considered and the chosen data analysis tehcniques are described as follows.

Main datasets

In the context of the National research project Pa.S.S.S., the City of Bari was chosen as field for road safety studies (public authorities involved in the Pa.S.S.S. project are the Ministry of Infrastructures and Transport and the City of Bari). Hence, datasets are related to the City of Bari (area of about 100 km^2, with 300,000 inhabitants).

The crash dataset for the City of Bari was provided by ASSET (Puglia Strategic Regional Agency for the Eco-Sustainable Area Development), which manage local crash data together with ISTAT (National Institute of Statistics). Crash data refer to the most recent five years at the beginning of the project, period: 2012-2016. They are crashes with at least one vehicle and one injured person involved (no property-damage only crashes), provided with exact localization (latitude and longitude GPS position), and several other information. Data about vehicles and persons involved, generic information (e.g., date, hour), crash (i.e., types and dynamics) and road-related variables are included in the crash dataset. For some crashes, it was not possible to precisely localize them due to inaccuracies in the positioning data.

Traffic data are essential for developing SPFs. They were initially provided by the City of Bari for the main road urban network. Due to the large extension of the urban road network, it was necessary to circumscribe an area of interest within the City. The main urban area was firstly selected (within the dashed line in Fig. 1) and the main interconnected urban network was defined within it, composed of arterial, collector and some high-level local roads. Available traffic data were coupled with crash data on the selected network. After, traffic counts were manually conducted to fill gaps in data obtained or anyway to check for inconsistencies due to old traffic volumes, especially on segments affected by new roads openings. At least one weekday peak hour traffic volume was obtained for the road sites used in this study, subsequently converted into average daily volumes through conversion factors.

Samples of sites

The selected network was further divided into segments and intersections. Crashes were then linked to each segment and intersection identified for this research. Some crashes have occurred in transition zones between segments and intersections (e.g., at pedestrian crossings near the intersection, but physically located on the adjacent road segment). In that case, no predefined rules (e.g., distance-based) were used to assign crashes to segments or intersections. Based on the crash localization and description (mainly crash type and dynamics), "transition" crashes were assigned to either segments or intersections. Traffic volumes were assigned to intersections by differentiating volumes on the main and the secondary entering roads. Segments were further divided into “homogeneous” segments, by considering geometric or traffic control differences within each segment (e.g., the introduction of another lane on the segment after some meters from the intersection would lead to divide the segment into two different homogeneous segments). Among the selected road network, a total number of 500 road “sites” were investigated: 325 homogeneous segments and 122 intersections.

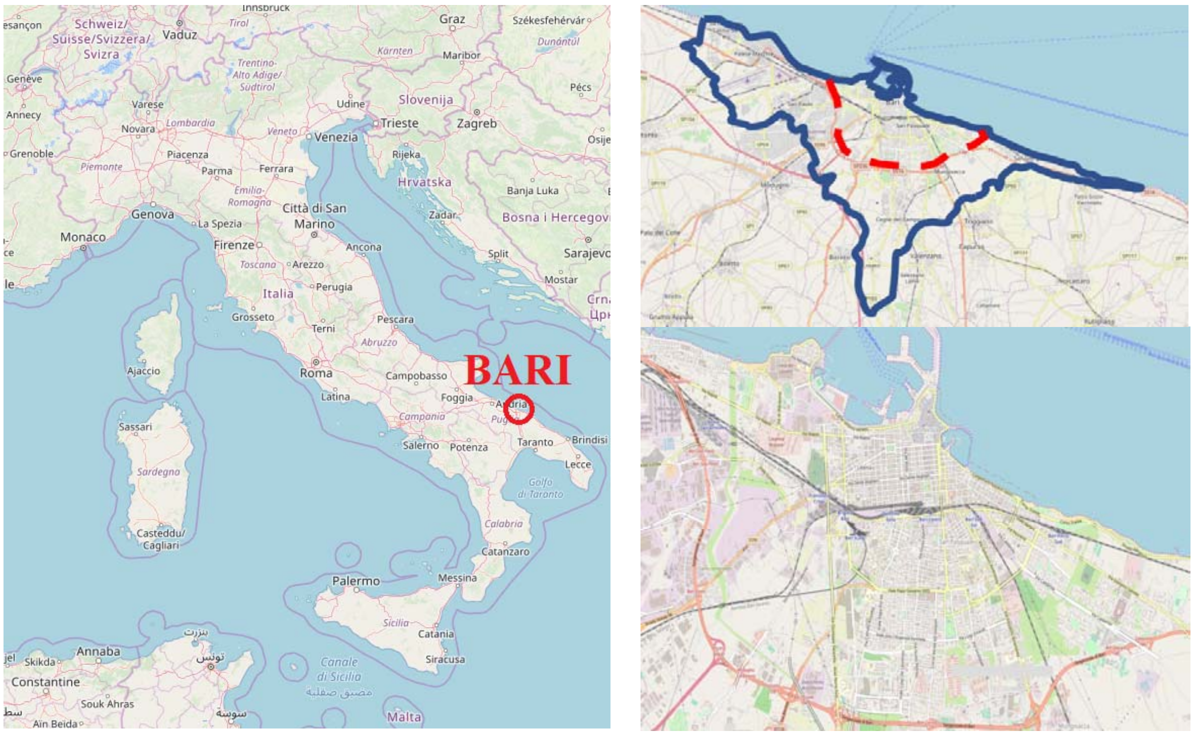

Figure 1 Localization of the city of Bari (within Italian country on the left). Dashed lines representing the limit of the urban area studied on the upper right (the main urban center inside the City of Bari, zoomed on the lower right)

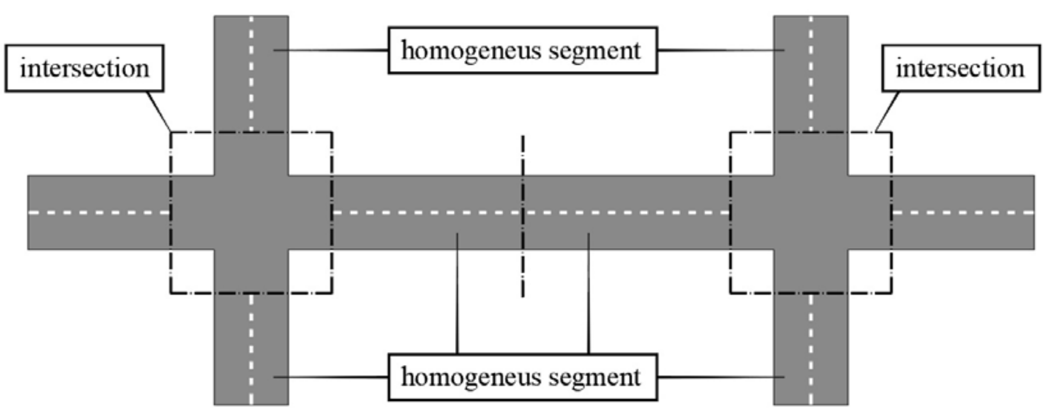

Figure 2 Division of the network into sites: intersections and homogeneous road segments

Crash predictors

Several crash predictors were taken into account in this study. Most information were derived from ad-hoc surveys and/or digital reconstructions of road sites. To allow further comparison with similar models developed in the same European area based on a previous literature review (see *11*), the main variables considered in (*1, 2*) were used here as well. Hence, data about the length of road sections, speed limits, road widths, number of minor roads/driveways, presence/type of parking, land use were collected for segments; while data about the number/width of lanes of intersecting roads, presence of medians on intersecting roads, presence/type of dedicated turning lanes, number of one-way legs were collected for intersections. In addition, other variables were considered such as sidewalks (*13*), presence of vertical signs on minor roads/driveways, maintenance of pavements, road markings, cycle paths, bus stops, reserved lanes (i.e., for taxi, buses), sight distance at intersections.

The selected variables are listed in Table 1. Some of them were categorical, while others were numerical. Categories in which variables were divided are indicated next to each variable, where relevant. Descriptive statistics (i.e., mean, standard deviation, minimum and maximum values, counts and percentages among the sample) are reported for each crash predictor. The computation of descriptive statistics is divided for segments and intersections and also for the different types of segments (i.e., one-way/two-way segments) and intersections (i.e., three-legged/four-legged intersections).

TABLE 1 Descriptive statistics of crash predictors for segments and intersections

Variables	**Modalities**	**Descriptive statistics**					
		Homogeneous Segments (n = 320*)		**One-way hom. segments (n = 190)**		**Two-way hom. segments (n = 130)**	
		Mean (st. dev.)/ Counts	Min-max/ Percentages	Mean (st. dev.)/ Counts	Min-max/ Percentages	Mean (st. dev.)/ Counts	Min-max/ Percentages
Fatal+Injury (F+I) crashes	Numerical	689	-	283	-	406	-
F+I crashes/segment/year	Numerical	0.43 (0.86)	(0-6)	0.30 (0.68)	(0-5)	0.63 (1.03)	(0-6)
One-way	0 – No (two-way) 1 – Yes	0 – 130 1 – 190	0 – 0.41 1 – 0.59	0 – 0 1 – 190	0 – 0.00 1 – 1.00	0 – 130 1 – 0	0 – 1.00 1 – 0.00
One-lane	0 –No (multilane) 1 – Yes	0 – 226 1 – 94	0 – 0.71 1 – 0.29	0 – 96 1 – 94	0 – 0.51 1 – 0.49	0 – 130 1 – 0	0 – 1.00 1 – 0.00
AADT	Numerical	8637.2 (5889.3)	300.0-29960.0	6228.8 (4569.5)	300.0-22600.0	12157.1 (5843.5)	1160.0-29960.0
Length (m)	Numerical	138.3 (183.3)	30.0-1510.0	124.1 (175.4)	30.0-1510.0	159.1 (193.0)	30.0-1415.0
Speed Limit = 50 km/h	0 – Yes 1 – No (> 50 km/h)	0 – 309 1 – 11	0 – 0.97 1 – 0.03	0 – 190 1 – 0	0 – 1.00 1 – 0.00	0 – 119 1 – 11	0 – 0.92 1 – 0.08
Road width (m)	Numerical	10.5 (3.0)	4.1-28.0	9.9 (2.6)	5.5-20.0	11.3 (3.5)	4.1-28.0
Type of lanes	0 – 1 lane 1 – > 1 lane (up to 3) 2 – 1+1 lane 3 – > 1+1 lanes (up to 3+3)	0 – 116 1 – 74 2 – 89 3 – 41	0 – 0.36 1 – 0.23 2 – 0.28 3 – 0.13	0 – 116 1 – 74 2 – 0 3 – 0	0 – 0.61 1 – 0.39 2 – 0.00 3 – 0.00	0 – 0 1 – 0 2 – 89 3 – 41	0 – 0.00 1 – 0.00 2 – 0.68 3 – 0.32

Density of minor roads/ driveways per km	Numerical	26.9 (26.0)	0.0-143.0	22.4 (22.8)	0.0-142.9	33.3 (28.9)	0.0-143.0
Vertical signs on minor roads/ driveways	0 – No 1 – Yes	0 – 256 1 – 64	0 – 0.80 1 – 0.20	0 – 158 1 – 32	0 – 0.84 1 – 0.16	0 – 98 1 – 32	0 – 0.75 1 – 0.25
Bad maintenance of pavements	0 – No 1 – Yes	0 – 71 1 – 249	0 – 0.22 1 – 0.78	0 – 49 1 – 141	0 – 0.26 1 – 0.74	0 – 22 1 – 108	0 – 0.17 1 – 0.83
Road markings	0 – No/Partially absent or illegible 1 – Illegible 2 – Yes	0 – 60 1 – 124 2 – 136	0 – 0.19 1 – 0.39 2 – 0.42	0 – 24 1 – 82 2 – 84	0 – 0.13 1 – 0.43 2 – 0.44	0 – 36 1 – 42 2 – 52	0 – 0.28 1 – 0.32 2 – 0.40
Parking type	0 – Prohibited 1 – At one side 2 – At both sides 3 – Mixed	0 – 85 1 – 109 2 – 119 3 – 7	0 – 0.27 1 – 0.34 2 – 0.37 3 – 0.02	0 – 26 1 – 74 2 – 87 3 – 3	0 – 0.14 1 – 0.39 2 – 0.46 3 – 0.02	0 – 59 1 – 35 2 – 32 3 – 4	0 – 0.45 1 – 0.27 2 – 0.25 3 – 0.03
Cycle paths	0 – No 1 – Yes	0 – 291 1 – 29	0 – 0.91 1 – 0.09	0 – 163 1 – 27	0 – 0.86 1 – 0.14	0 – 128 1 – 2	0 – 0.98 1 – 0.02
Sidewalks	0 – No 1 – Yes	0 – 31 1 – 289	0 – 0.10 1 – 0.90	0 – 4 1 – 186	0 – 0.02 1 – 0.98	0 – 27 1 – 103	0 – 0.21 1 – 0.79
Median	0 – No 1 – Yes	0 – 244 1 – 76	0 – 0.76 1 – 0.24	0 – 114 1 – 76	0 – 0.60 1 – 0.40	0 – 130 1 – 0	0 – 1.00 1 – 0.00
Bus stop	0 – No 1 – Yes	0 – 227 1 – 93	0 – 0.71 1 – 0.29	0 – 143 1 – 47	0 – 0.75 1 – 0.25	0 – 84 1 – 46	0 – 0.65 1 – 0.35
Bus/taxi lane	0 – No 1 – Yes	0 – 284 1 – 36	0 – 0.89 1 – 0.11	0 – 154 1 – 36	0 – 0.81 1 – 0.19	0 – 130 1 – 0	0 – 1.00 1 – 0.00
Land use	0 – Residential 1 – Mainly commercial 2 – Other	0 – 115 1 – 150 2 – 55	0 – 0.36 1 – 0.47 2 – 0.17	0 – 71 1 – 110 2 – 9	0 – 0.37 1 – 0.58 2 – 0.05	0 – 44 1 – 40 2 – 46	0 – 0.34 1 – 0.31 2 – 0.35
		Intersections (n = 120*)		**Three-legged intersections (n = 48)**		**Four-legged intersections (n = 72^)**	
		Mean (st. dev.)/ Counts	Min-max/ Percentages	Mean (st. dev.)/ Counts	Min-max/ Percentages	Mean (st. dev.)/ Counts	Min-max/ Percentages
Fatal+Injury (F+I) crashes	Numerical	906	-	232	-	674	-
F+I crashes/intersection/year	Numerical	1.51 (1.84)	0-15	0.97 (1.41)	0-15	1.87 (1.99)	0-12
Traffic lights	0 – No 1 – Yes	0 – 59 1 – 61	0 – 0.49 1 – 0.51	0 – 38 1 – 10	0 – 0.79 1 – 0.21	0 – 21 1 – 51	0 – 0.29 1 – 0.71
Type = 4 legs	0 – No 1 – Yes	0 – 48 1 – 72	0 – 0.40 1 – 0.60	0 – 48 1 – 0	0 – 1.00 1 – 0.00	0 – 0 1 – 72	0 – 0.00 1 – 1.00
Main AADT	Numerical	12192.4 (5313.2)	2475.0-23680.0	12588.4 (5350.8)	3550.0-22340.0	11928.4 (5309.1)	2475.0-23680.0
Secondary AADT	Numerical	5084.1 (4271.5)	300.0-17880.0	4008.3 (3897.4)	300.0-18520.0	5801.4 (4384.2)	345.0-17880.0
Total AADT	Numerical	17276.5 (7981.8)	4175.0-41560.0	16596.6 (7018.3)	5420.0-32300.0	17729.7 (8582.2)	4175.0-41560.0
Main AADT/ Total AADT	Numerical	0.53 (0.03)	0.50-0.62	0.55 (0.03)	0.50-0.62	0.53 (0.02)	0.50-0.61
Lane balance	0 – No 1 – Yes	0 – 81 1 – 39	0 – 0.67 1 – 0.33	0 – 38 1 – 10	0 – 0.79 1 – 0.21	0 – 43 1 – 29	0 – 0.60 1 – 0.40
Median on the main road	0 – No 1 – Yes	0 – 100 1 – 20	0 – 0.83 1 – 0.17	0 – 48 1 – 0	0 – 1.00 1 – 0.00	0 – 52 1 – 20	0 – 0.72 1 – 0.28
Median on the secondary road	0 – No 1 – Yes	0 – 107 1 – 13	0 – 0.89 1 – 0.11	0 – 43 1 – 5	0 – 0.90 1 – 0.10	0 – 64 1 – 8	0 – 0.89 1 – 0.11

Entering lanes on the main road	Numerical	1.5 (1.1)	0.0-5.0	1.1 (1.0)	0.0-4.0	1.8 (1.1)	0.0-5.0
Mean lane width at intersections (m)	Numerical	4.5 (1.1)	2.8 (8.7)	4.3 (1.0)	2.8-6.6	4.6 (1.2)	2.8-8.7
Worst sight distance (m)	Numerical	31.2 (26.0)	0.0-139.5	30.0 (28.1)	0.0-139.5	32.0 (24.6)	1.6-100.0
One-way legs	Numerical	1.7 (1.5)	0.0-4.0	1.4 (1.1)	0.0-3.0	2.0 (1.6)	0.0-4.0
Specialized turning lane	0 – No 1 – Yes	0 – 88 1 – 32	0 – 0.73 1 – 0.27	0 – 43 1 – 5	0 – 0.90 1 – 0.10	0 – 45 1 – 27	0 – 0.63 1 – 0.37
Traffic control	0 – No 1 – Give-way/Stop 2 – Traffic lights 3 – Traffic lights with dedicated turning allowed	0 – 22 1 – 37 2 – 35 3 – 26	0 – 0.18 1 – 0.31 2 – 0.29 3 – 0.22	0 – 16 1 – 22 2 – 7 3 – 3	0 – 0.33 1 – 0.46 2 – 0.15 3 – 0.06	0 – 6 1 – 15 2 – 28 3 – 23	0 – 0.08 1 – 0.21 2 – 0.39 3 – 0.32
Bad maintenance of pavements	0 – No 1 – Yes	0 – 18 1 – 102	0 – 0.15 1 – 0.85	0 – 9 1 – 39	0 – 0.19 1 – 0.81	0 – 9 1 – 63	0 – 0.13 1 – 0.87
Cycle path crossing	0 – No 1 – Yes	0 – 103 1 – 17	0 – 0.86 1 – 0.14	0 – 36 1 – 12	0 – 0.75 1 – 0.25	0 – 67 1 – 5	0 – 0.93 1 – 0.07
Sidewalks	0 – No 1 – Yes (both sides)	0 – 8 1 – 112	0 – 0.07 1 – 0.93	0 – 3 1 – 45	0 – 0.06 1 – 0.94	0 – 5 1 – 67	0 – 0.07 1 – 0.93
Bus stops	0 – No 1 – Yes	0 – 111 1 – 9	0 – 0.93 1 – 0.07	0 – 46 1 – 2	0 – 0.96 1 – 0.04	0 – 65 1 – 7	0 – 0.90 1 – 0.10
Bus/taxi lanes on intersecting roads	0 – No 1 – Yes	0 – 92 1 – 28	0 – 0.77 1 – 0.23	0 – 43 1 – 5	0 – 0.90 1 – 0.10	0 – 49 1 – 23	0 – 0.68 1 – 0.32

[*]Note: Five segments were discharged from the initial dataset due to segment length minor than 30 m, which were deemed as irrelevant for safety modelling purposes. Two segments were discharged from the initial dataset due to possible errors in the counts of traffic volumes, leading to unrealistic data.
^These intersections include one five-legs intersection.

Data analysis techniques

The development of SPFs for linking crash frequencies to predictors require the application of statistical techinques. In particular, NB count data models (General Linear Models -GLM- assuming an underlying negative binomial distribution of the errors) are widely used for this aim. In fact, they can account for the over-dispersion of crash data (*15*). They were used in similar studies in which urban SPFs were developed in previous studies (2, 4-5). NB models were developed in the R software environment, by using the library MASS (*16*). The general model structure used is reported as follows:

$$N_{SPF} = e^{\beta 0} * AADT^{\beta 1} * L^{\beta 2} * e^{\sum_{i=3}^{n} \beta i X i} \tag{1}$$

Where:

AADT = Annual Average Daily Traffic (single value for segments or divided into main and secondary volume for intersections, or also treated as total volume in the intersection);

L = segment length (in meters, not present for intersections);

X_i = other crash predictors (numerical or categorical, in case of categorical variables they are transformed into binary dummy variables with modalities 0 and 1, 0 is the reference modality);

β_i = estimate of the coefficients associated to each crash predictor through maximum likelihood estimation (β_0 is the estimate for the intercept).

One of the research questions of this study concerned the most appropriate way of disaggregating segments and intersections into more detailed subsets for crash prediction purposes. For this aim, the modelling stage was further divided into different steps:

- Running preliminary models for the whole dataset of both segments and intersections with the aim of identifying variables useful for further classifications (i.e., for segments: one-way/two-way, one-lane/multilane, for intersections: unsignalized/signalized, three-legged/four-legged);
- Selecting two sub-categories for each family of sites (segments and intersections) based on results from preliminary models;
- Running models for each of the sub-categories of sites identified.

For each model run, starting from the initial full set of variables, the model having the least number of all significant variables included was selected, by subsequently trying different combinations of variables. Besides of being used for future local predictions, this study has an exploratory purpose for testing the significance of some directly observed variables and the effectiveness of considering different subsets of homogeneous road sites. The choice of disaggregating the dataset into different subsets for research purposes result in reducing the size of the samples. The chosen level of significance was then set to $p = 0.10$, given the exploratory purposes and the limited dataset (similar to previous studies, see e.g., *2*).

Results from each model obtained were compared to the corresponding null models, to assess the overall significance of fits, by means of likelihood ratio tests (5 % significance level). The same test was used for comparing final models with the full models, to assess if the reduced number of variables has produced a significant improvement or not (null hypothesis: the model with less variables has better prediction capabilities than the model with more variables, full model in this case). The Akaike Information Criterion (AIC) was also used to comparatively assess different models. The Nagelkerke R^2 (variable in the range: 0-1) was the main goodness-of-fit measure used to assess each model.

RESULTS

Results from the estimated models are shown in this section. The presentation of results is divided according to the type of road elements investigated (segments and intersections). The interpretation of the results obtained as well as the influence of the significant predictors will be discussed in detail in the discussion section further below.

Predictive models for segments

A model was firstly developed by considering the whole dataset of segments. The list of crash predictors for segments reported in Table 1 were taken into account altogether. However, some of them were clearly correlated, such as the variables representing the number of lanes and the one-way variable. Hence, different tests were conducted by trying different combinations of variables and avoiding to include together correlated predictors.

The main interest in this stage was to understand if it was possible to further differentiate the initial dataset into different subsets. Among the different attempts, the variable "Type of lanes" seems to have more explanatory power than the others. However, it naturally includes the differences between one-way and two-way segments (reference type 0 and type 1 are for one-way segments, types 2 and 3 are for two-way undivided segments). On the other hand, the variable "One-lane" is never found as a statistically significant predictor. For this reason, the dataset was divided into two subsets: one-way and two-way segments, rather than one-lane and multilane segments. Note that each segment divided by medians was firstly divided into two one-way segments (one for each direction). This was made since: a) traffic counts were available for each direction, b) each homogeneous segment for which a median was present, is defined so that the median is not crossable for all the homogeneous segment length, c) medians of the sites considered are often large and organized with sidewalks. Hence, the possible division: undivided/divided segments was not taken into account.

The model developed for all the segments is reported in Table 2. Clearly, crash frequencies increase with traffic volumes and segment length. Two-way segments are associated with increased crash frequencies with respect to one-way segments, as based on the variable "Type of lanes", while it seems that no differences exist between one-lane and multilane one-way segments in terms of safety. The presence of vertical signs on driveways or intersecting minor roads is associated with an increase in crashes, as well as parking on both sides (with respect to no parking allowed). Bad maintenance of pavements is associated with a decrease in the number of crashes. Some predictors are statistically significant only at the 10 % level: parking type at one side (related to crash increasing) and presence of cycle paths close to segments (related to crash increasing as well). However, it should be noted that the included predictors can only explain a limited part of the crash frequencies, as based on the Nagelkerke R^2 value (which is not drastically decreased with respect to the full model).

TABLE 2 Selected models for segments

Predictor	Estimate	Std. Error	*z*- value	*p*-value
All segments				
(Intercept)	-6.054	0.728	-8.318	< 0.001
Ln(AADT)	**0.375**	**0.084**	**4.449**	**< 0.001**
Ln(L)	**0.292**	**0.060**	**4.907**	**< 0.001**
Type of lanes 1: >1 (reference: 1 lane)	0.079	0.167	0.470	0.638
Type of lanes 2: 1+1 (ref.: 1 lane)	**0.524**	**0.151**	**3.471**	**< 0.001**
Type of lanes 3: > 1+1 (ref.: 1 lane)	**0.707**	**0.181**	**3.906**	**< 0.001**
Vertical signs on driveways/minor roads (Yes - ref.: No)	**0.652**	**0.110**	**5.943**	**< 0.001**
Bad maintenance of pavements (Yes - ref.: No)	**-0.268**	**0.114**	**-2.348**	**0.019**
Parking type 1: At one side (ref.: No parking)	*0.218*	*0.131*	*1.668*	*0.095*
Parking type 2: At both sides (ref.: No parking)	**0.258**	**0.126**	**2.053**	**0.040**
Parking type 3: Mixed (ref.: No parking)	**-1.092**	**0.556**	**-1.964**	**0.050**
Cycle paths (Yes – ref.: No)	*0.308*	*0.182*	*1.694*	*0.090*
Goodness-of-fit measures				
LRT (reference: null model) = 200.96 (df = 11), p < 0.001; LRT (reference: full model) = 6.64 (df = 11), p = 0.828				
Nagelkerke R^2 = 0.158, AIC = 2609.9				
Over-dispersion parameter (theta) = 1.08 (std. error: 0.17)				
One-way segments				
(Intercept)	-8.794	0.961	-9.151	< 0.001

Ln(AADT)	**0.695**	**0.108**	**6.434**	**< 0.001**
Ln(L)	**0.333**	**0.095**	**3.505**	**< 0.001**
Driveways/minor roads per km	**0.010**	**0.003**	**3.163**	**0.002**
Vertical signs on accesses/minor roads (Yes - ref.: No)	**0.477**	**0.169**	**2.825**	**0.005**
Bad maintenance of pavements (Yes - ref.: No)	**-0.497**	**0.159**	**-3.137**	**0.002**
Goodness-of-fit measures				
LRT (reference: null model) = 101.68 (df = 5), $p < 0.001$; LRT (reference: full model) = 9.77 (df = 10), $p = 0.460$				
Nagelkerke $R^2 = 0.153$, AIC = 1226.8				
Over-dispersion parameter (theta) = 1.01 (std. error: 0.26)				
Two-way segments				
(Intercept)	-3.764	1.122	-3.356	< 0.001
Ln(AADT)	0.134	0.117	1.140	0.254
Ln(L)	**0.310**	**0.079**	**3.944**	**< 0.001**
Vertical signs on accesses/minor roads (Yes - ref.: No)	**0.432**	**0.141**	**3.074**	**0.002**
Markings 1: Illegible (ref.: No/partially absent/illegible)	*0.317*	*0.177*	*1.791*	*0.073*
Markings 2: Yes (ref.: No/partially absent/illegible)	*0.309*	*0.175*	*1.767*	*0.077*
Parking type 1: At one side (ref.: No parking)	*0.287*	*0.156*	*1.836*	*0.066*
Parking type 2: At both sides (ref.: No parking)	**0.441**	**0.152**	**2.903**	**0.004**
Parking type 3: Mixed (ref.: No parking)	-0.973	0.640	-1.521	0.128
Goodness-of-fit measures				
LRT (reference: null model) = 57.69 (df = 7), $p < 0.001$; LRT (reference: full model) = 5.63 (df = 7), $p = 0.584$				
Nagelkerke $R^2 = 0.102$, AIC = 1369.2				
Over-dispersion parameter (theta) = 1.32 (std. error: 0.29)				

When differentiating into one-way and two-way segments, some of the significant predictors are confirmed, while others are highlighted as well. Traffic volumes and segment lengths are confirmed as crash predictors in the one-way segment model. As the number of driveways and minor roads on the one-way segment increases, the crash frequency increases. Vertical signs on minor roads/driveways are confirmed as associated with increasing crashes for both one-way and two-way segments. The same negative coefficient estimate is confirmed for the bad maintenance of pavements in the one-way case. Parking especially at both sides is associated to increasing crashes for two-way segments with respect to prohibited parking. The presence of visible marking is associated with increasing crashes, although marginically significant at the 5 % level. Note that the coefficient of segment length is positive and significant, while that for traffic volume is positive but not significant for two-way segments (even if included in the final model selected). However, the model with traffic volume included was anyway selected among other possibilities (its coefficient is statistically significant with $p < 0.10$ if considered as the only predictor for crashes without covariates), also to avoid further worsening the limited model fit.

Predictive models for intersections

A model was firstly developed by considering the whole dataset of intersections. The list of crash predictors for intersections reported in Table 1 were taken into account altogether. However, also in this case, some of them are clearly correlated, such as the variable signalized/unsignalized and the different types of intersection control. To some extend, also the presence of specialized turning lane and the type of intersection (three/four-legged) may be correlated among them and with the other predictors mentioned before. In fact, almost all three-legged intersections are unsignalized and almost all specialized lanes are in four-legged intersections. Hence, also in this case, different tests were conducted by trying different combinations of variables, by avoiding to include together correlated predictors.

The main interest in this stage was to understand if it is possible to further differentiate the initial dataset into different subsets. In this case, it is possible to select two promising models for the whole intersections (see Table 3: "All intersections" and "All intersections (alternative)" models). Thus, considering the predictors which may help in individuating possible subsets (mainly: unsignalized/signalized and three-legged/four-legged intersections) and the different attempts performed, the first model selected in Table 3 indicates the number of legs as an important explanatory variable. However, when trying to exclude all other possible correlated variables (turning lanes, number of legs, type of intersection control), the variable signalized/unsignalized assumes a notable importance in an alternative model. Hence, based on this, specific models were developed for two pairs of subsets: three-legged and four-legged intersections, signalized and unsignalized intersections. The consideration of the signalized/unsignalized subsets (which could have been excluded at a first attempt) can be important for the practical models use with predictive purposes. Moreover, another choice was made between considering together the main and secondary traffic volumes in the model, or the total volume and the ratio of the main to the total volume. The second alternative has generally led to a better goodness-of-fit of the models, as based on the attempts made.

Both overall models for intersections include the total traffic volume and the main-to-total AADT ratio as significant variables. As the total volume increases, the crash frequency increases. However, as the main-to-total AADT ratio increases, the crash frequency decreases. Putting these two together show that crashes increases at a decresing rate as the total AADT increases. Four-legged intersections seem comparatively less safe than three-legged intersections and signalized interections slightly less safe than the unsignalized ones. The negative coefficient for bad maintenance of pavements is confirmed for intersections too. The presence of specialized turning lanes and cycle paths are associated to crash increasing in the first model in Table 3. The alternative model for all intersections highlights the importance (on crash increasing) of critical sight distance instead. Predictors for intersections can explain the crashes variability better than in the segments case, as based on the Nagelkerke R^2 values.

TABLE 3 Selected models for intersections

Predictor	Estimate	Std. Error	z- value	*p*-value
All intersections				
(Intercept)	0.930	1.524	0.610	0.542
Ln(Total AADT)	**0.285**	**0.106**	**2.675**	**0.007**
Main AADT/Total AADT	**-6.870**	**2.021**	**-3.399**	**0.001**
Type of intersection: Four-legged (reference: Three-legged)	**0.443**	**0.105**	**4.210**	**< 0.001**
Critical sight distance	**0.005**	**0.002**	**2.762**	**0.006**
Specialized turning lane (Yes – ref.: no)	**0.496**	**0.114**	**4.362**	**< 0.001**
Bad maintenance of pavements (Yes - ref.: No)	**-0.397**	**0.115**	**-3.444**	**< 0.001**
Cycle path crossing (Yes – ref.: No)	**0.327**	**0.122**	**2.677**	**0.007**
Goodness-of-fit measures				
LRT (reference: null model) = 5.63 (df = 12), p = 0.934; LRT (reference: full model) = 161.11 (df = 12), p < 0.001				
Nagelkerke R^2 = 0.283, AIC = 1877.6				

Over-dispersion parameter (theta) = 2.84 (std. error: 0.52)				
Three-legged intersections				
(Intercept)	4.696	2.895	1.609	0.108
Ln(Total AADT)	*0.366*	*0.220*	*1.661*	*0.097*
Main AADT/Total AADT	**-16.376**	**3.711**	**-4.413**	**<0.001**
Number of entering lanes on the main road	*-0.211*	*0.119*	*-1.768*	*0.077*
Critical sight distance	**0.006**	**0.003**	**2.009**	**<0.001**
Specialized turning lane (Yes – ref.: no)	**0.927**	**0.228**	**4.061**	**<0.001**
*Traffic control 1: give-way/stop (ref.: no control)**	*0.359*	*0.214*	*1.675*	*0.094*
Traffic control 2: traffic lights (ref.: no control)*	0.377	0.289	1.306	0.192
*Traffic control 3: main+turning lights (ref.: no control)**	*0.556*	*0.313*	*1.777*	*0.076*
Cycle path crossing (Yes – ref.: No)	**0.900**	**0.235**	**3.830**	**<0.001**
Goodness-of-fit measures				
LRT (reference: null model) = 63.39 (df = 9), p < 0.001; LRT (reference: full model) = 5.79 (df = 9), p = 0.761				
Nagelkerke R^2 = 0.292, AIC = 610.4				
Over-dispersion parameter (theta) = 5.26 (std. error: 2.75)				
Four-legged intersections				
(Intercept)	-0.571	1.823	-0.313	0.754
Ln(AADT)	**0.381**	**0.131**	**2.905**	**0.004**
Main AADT/Total AADT	*-4.623*	*2.620*	*-1.764*	*0.078*
Critical sight distance	**0.005**	**0.002**	**2.263**	**0.024**
Specialized turning lane (Yes – ref.: no)	**0.418**	**0.140**	**2.981**	**0.003**
Bad maintenance of pavements (Yes - ref.: No)	**-0.529**	**0.152**	**-3.476**	**0.001**
LRT (reference: null model) = 77.98 (df = 5), p < 0.001; LRT (reference: full model) = 8.92 (df = 11), p = 0.630				
Nagelkerke R^2 = 0.225, AIC = 1266.7				
Over-dispersion parameter (theta) = 2.71 (std. error: 0.56)				
All intersections (alternative)				
(Intercept)	1.713	1.513	1.132	0.257
Ln(Total AADT)	**0.371**	**0.103**	**3.613**	**< 0.001**
Main AADT/Total AADT	**-9.502**	**2.161**	**-4.397**	**< 0.001**
Median on the main road (Yes – reference: No)	*0.226*	*0.117*	*1.930*	*0.054*
Critical sight distance	**0.004**	**0.002**	**2.359**	**0.018**
Type of intersection (Signalized – ref.: Unsignalized)	**0.268**	**0.104**	**2.580**	**0.010**
Bad maintenance of pavements (Yes - ref.: No)	**-0.321**	**0.118**	**-2.730**	**0.006**
Cycle path crossing (Yes - ref.: No)	**0.253**	**0.123**	**2.058**	**0.040**
Goodness-of-fit measures				
LRT (reference: null model) = 124.59 (df = 7), p < 0.001; LRT (reference: full model) = 8.29 (df = 8), p = 0.406				
Nagelkerke R^2 = 0.222, AIC = 1914.1				
Over-dispersion parameter (theta) = 2.40 (std. error: 0.41)				
Signalized intersections				
(Intercept)	2.390	2.293	1.043	0.297
Ln(Total AADT)	**0.407**	**0.142**	**2.863**	**0.004**
Main AADT/Total AADT	**-11.302**	**3.586**	**-3.152**	**0.002**
Type of intersection: Four-legged (reference: Three-legged)	**0.452**	**0.182**	**2.479**	**0.013**
Critical sight distance	**0.004**	**0.002**	**1.986**	**0.047**
Specialized turning lane (Yes – ref.: no)	**0.307**	**0.148**	**2.076**	**0.038**
Bad maintenance of pavements (Yes - ref.: No)	**-0.630**	**0.148**	**-4.243**	**< 0.001**
Presence of bus stops (Yes – ref.: No)	**-0.914**	**0.357**	**-2.559**	**0.010**
Goodness-of-fit measures				
LRT (reference: null model) = 95.60 (df = 7), p < 0.001; LRT (reference: full model) = 4.39 (df = 9), p = 0.884				

Nagelkerke R^2 = 0.320, AIC = 1078.2				
Over-dispersion parameter (theta) = 3.08 (std. error: 0.71)				
Unsignalized intersections				
(Intercept)	-2.604	2.185	-1.192	0.233
Ln(AADT)	**0.759**	**0.163**	**6.343**	**< 0.001**
Main AADT/Total AADT	**-7.868**	**0.173**	**4.389**	**0.001**
Type of intersection: Four-legged (reference: Three-legged)	**1.036**	**2.406**	**-3.270**	**< 0.001**
Median on the main road (Yes – reference: No)	**-0.998**	**0.329**	**-3.031**	**0.002**
Specialized turning lane (Yes – ref.: No)	**0.516**	**0.204**	**2.523**	**0.012**
Cycle path crossing (Yes - ref.: No)	**0.997**	**0.193**	**5.156**	**< 0.001**
Presence of sidewalks (Yes - ref.: No)	**-1.008**	**0.331**	**-3.042**	**0.002**
LRT (reference: null model) = 65.32 (df = 7), p < 0.001; LRT (reference: full model) = 10.46 (df = 11), p = 0.489				
Nagelkerke R^2 = 0.239, AIC = 769.6				
Over-dispersion parameter (theta) = 8.11 (std. error: 6.16)				

*Note: The variable traffic control device was further disaggregated into: no control, give-way, stop control for unsignalized intersections, and into: traffic lights, traffic lights with dedicated left turning lights, traffic lights with dedicated right turning lights for signalized intersections.

In the three-legged model, sight distance, turning lanes and cycle path crossing are confirmed as predictors (with similar coefficients). The presence of traffic lights does not seem to be influential with respect to the no-control condition (except for traffic lights with dedicated turning lights, which seem to be related to an increase in crashes). Moreover, it seems that as the entering lanes on the main road increase, the number of crashes decreases ($p < 0.10$). In the four-legged model, sight distance and turning lanes are confirmed as significant predictors, while cycle path crossings are not. The presence of traffic lights seems not influential at all, even if its effect may be partially included in the specialized turning lane variable. The bad maintenance of pavements is here related to crashes decreasing, such as for segments.

In the signalized model (having the highest R^2 value among the presented models), it is confirmed that four-legged intersections are comparatively less safe than three-legged intersections (such as in the unsignalized model). As the critical sight distance increases, there is an increase in the crash frequency for the signalized intersections (similar to models for three/four-legged intersections). Specialized turning lanes and bad maintenance of pavements are confirmed as crash predictors (the first with a positive coefficient and the second with a negative coefficient). The presence of bus stops close to signalized intersections appears here as a factor associated with an increased crash frequency. In the unsignalized model, some other predictors are highlighted, besides of those already mentioned. The presence of median on the main road is associated with a decrease in the number of crashes ($p < 0.10$), as well as the presence of sidewalks surrounding the intersection.

DISCUSSION

The results obtained from this work are discussed here in light of the research questions posed in the introductory section: a) finding the best possible subsets of urban segments and intersections according to local conditions; b) comparing results with previous relevant research to possibly highlight some specific local conditions or common trends; c) assessing some variables often not considered for predictions, mainly based on visual observations. The first two points were merged in the following discussion section, since discussing the different models selected for subsets of segments and intersections includes also discussing the significant predictor variables. The comparison of trends found for the different predictors was compared with some relevant

European studies in order to allow a consistent comparison, as based on a previous literature review (*11*). Very few detailed recent studies which have considered several predictor variables for road segments were found. On the other hand, more studies were found for the intersections, some of them differentiated for categories of intersections, as well as in this study.

Subsets of urban segments and intersections for safety modelling and associated predictors
The predictive models for urban segments developed in this study were stratified into models for one-way segments and two-way segments. In fact, significant differences were found between one-way segments and two-way segments (see overall model for segments in Table 2), differently than in a previous study (*1*), in which the variable one-way/two-way was not included in the final model for segments. Two-way segments result as less safe compared to one-way segments (through a factor of $e^{0.524} = 1.689$ for two-way two-lane segments and a factor of $e^{\cdot 0.707} = 2.028$ for two-way multilane segments, with respect to one-lane segments), other conditions being equal. On the other hand, significant differences between one-lane and multilane segments (either one-way or two-way travelled) were not highlighted. The variable "Type of lanes" was not included in the separated selected models for one-way and two-way segments. Hence, it seems that on one-way urban segments, the organization in one or more lanes does not seem influential on crash risk. However, it should be noted that the mean road width is about 10 m for one-way roads. Thus, even single-lane one-way segments may have very wide lanes, which could be anyway practically travelled on two lanes by operating vehicles (even if two lanes are not indicated by markings). Significant differences between two-way segments with one lane per direction and two-way segments with more lanes per direction were not highlighted as well in the dedicated model for two-way segments. Note however that in the overall model for segments, the type of lanes "3" (multilane two-way segments) seems slightly less safe than the type of lanes "2" (two-way two-lanes segments) other conditions being equal, with respect to the one-lane reference "0" condition ($e^{\text{Typeoflanes3}}/e^{\text{Typeoflanes2}} = e^{0.707}/e^{0.524} = 1.20$), coeherenly with what reported in the HSM (*6*).

On the other hand, the predictive models for urban intersections were stratified into models for signalized and unsignalized intersections and for three-legged and four-legged intersections. In fact, significant differences were found between both categories. Note that, as expected, four-legged intersections result as comparatively less safe than three-legged intersections: accidents increase by a factor of $e^{0.443} = 1.557$ in the case of four-legged intersections (other conditions being equal). Clearly, this can be attributed to the higher number of conflicts in four-legged intersections, potentially having more vehicular crossing flows than three-legged intersections. The effect of signals on intersections is less clear however. This is coherent with a previous study (*2*), in which different models were developed for three/four-legged intersections, while the traffic signal variable was not significant. However, very disaggregated subsets of intersections were considered in (*13*) for modelling purposes: three/four-legged no-control/stop-controllled, or four-legged signalized intersections. However, in this study, differences in predictors were noted by dividing the whole dataset in unsignalized and signalized intersections. Moreover, contrary to expectations, signalized intersections seem comparatively slightly less safe than unsignalized intersections (factor of $e^{0.268} = 1.307$), as it emerged from the alternative model in which other intersection-type related variables were excluded. The division into signalized and unsignalized intersections allows to highlight that four-legged intersections are comparatively less safe than three-legged intersections especially for unsignalized intersections, as expected (coefficient estimate > 1 for four-legged unsignalized intersections). Note also that the give-way control of three-legged intersections seems even less safe than no-control intersections, ceteris paribus, while other conditions can be hardly assessed. Moreover, the presence of specialized turning lanes (left/right) seems consistently negative for safety. This could be explained by: a) turning lanes being a surrogate measure for the importance of the intersection, thus intrinsically leading to more conflicts, b) drivers not respecting road rules and then causing additional conflicts when specialized turning lanes are present. This effect can be particularly more important for three-legged and unsignalized intersections, where however the presence of turning lanes is greatly limited. Similar negative effects on safety were previously already highlighted for the presence of right turn on the major road at both three/four-legged intersections (*2*). In another study (*12*), mixed results were found for left/right turning lanes according to the specific intersection type.

In any case, developing models for different subsets of urban intersections and segments is extremely useful for identifying predictors which are specifically related to some types of segments or intersections. For example, considering segments, the increasing number of minor roads on segments is generally associated to an increase in the number of crashes (*1, 12*). In this study, a first important difference was noted for the number of driveways and minor roads, which is significant only for one-way segments. One-way roads (especially if wide as in the considered dataset) may allow high speeds. Moreover, drivers should not care about other vehicles eventually crossing the travel direction from the other lane, such as it happens on two-way undivided roads. Another difference relates to the presence of parking, which is generally associated to crash increasing with respect to rarely/prohibited parking (*1*). In this study, the presence of parking (especially at both sides) is a significant crash predictor (related to an increase in crashes) only for two-way segments. Clearly, on two-way undivided roads, the possible conflicts due to parking on both sides are greater than on one-way roads. In this case, conflicts may even be more unexpected than in case of known minor roads and then drivers on two-way roads could not react in time, differently than at intersections with driveways/minor roads (which seem less influential on two-way segments). The coefficient estimate for the traffic volume is low: the relationship is far less than linear for both one-way and two-way segments ($\beta < 1$). However, the coefficient is very close to zero (not significant at the 10 % level) in case of two-way segments. This occurrence could be hardly generalized, and rather be attributed to the limited sample of two-way segments. However, it seems that the effect of urban congestion (with traffic increasing) is more detrimental for safety on one-way segments than on two-way segments. Note that speed limits, road width and land use were not included in the selected models for segments, differently than in (*1*). However, speed limits are almost always equal to 50 km/h on the selected network and land use is not largely variable due to the selection of sites within the central city area (see Fig. 1).

For what concerns intersections, the coefficients estimated for traffic volumes (and main-to-total ratio) are similar. However, note that when the ratio main-to-total AADT increases by 10 % (that is when the secondary-to-total volume decreases by 10 %), crashes increase with a factor of about 5 for three-legged intersections (in this case, it is statistically significant at the 5 % level) and about 1.5 for four-legged intersections, other conditions being equal. Hence, it could be stated that, especially for three-legged intersections, a severe unbalance between main and secondary volumes at intersections may generate an increase in the number of crashes, coherently with results from a previous study (*2*). In this case, vehicles on the main road may not decrease their speeds in presence of an intersection with a limited number of vehicles entering in the main flow. Moreover, in case of signalized intersections (where the AADT ratio coefficient is still high and significant), drivers may even ignore red lights for cases when few vehicles are entering from the secondary road. It is worth noting that a traditional crash predictor such as the presence of medians on the main intersecting road emerges as significant ($p < 0.10$), but only for unsignalized intersections: in this case their presence may be positive for safety. This is in line with previous results specific for three-legged intersections (*2*), in particular stop-controlled (*12*). In fact, the presence of medians may help in physically channelize the intersecting traffic flow. Note also that road markings were previously found as negative for safety for three-legged stop-control intersections (*12*), such as in this study for two-way segments. Other predictors highlighted in previous studies were not included here in the finally selected models, such as the lane balance, the number of intersecting one-way legs (*2*), the presence of two-way on the major intersecting road and the lane width at intersections (*12*).

Assessment of additional variables for safety predictions

Some additional variables which are usually not often considered for safety predictions were included in this study. For example, the presence of vertical signs on minor roads/driveways, maintenance of pavements, road markings, cycle paths, bus stops, reserved lanes (i.e., for taxi, buses), sight distance at intersections, were considered. Some of them were actually included in the selected models presented in the results section. However, most of those present some unexpected trends, which may seem surprising at a first glance. For example, the presence of vertical signs on driveways/minor roads intersecting with the segments seem to provide a negative effect for safety. However, it should be noted that the number of driveways/minor roads provided with vertical signs is scarce among the sample of sites (see Table 1). Hence, it is likely that this variable could be a surrogate measure for the importance of the driveway/minor road, i.e., considering the unlikely event that very-low volume driveways may have any vertical sign on it. Clearly, this means that not all driveways/minor roads may have the same effect on safety.

Other additional variables which are worth to mention are those related to maintenance: markings and pavement conditions. In fact, it seems counterintuitive that deteriorated pavements may be positive for safety (as found for one-way segments, four-legged and signalized intersections), as well as well maintained and visible markings (as found for two-way segments). This could be explained by drivers being more cautious and driving at lower speeds on bad maintained pavements. However, the more likely suggested interpretation is associated to the temporal displacement between the visual observations mainly conducted during 2018 and the crashes in the dataset referred to the period 2012-2016. Hence, it is most likely that pavements (and then markings) in good conditions in 2018 could have been resurfaced in the last years and vice versa. Thus, considering the service life span (especially of pavements) is most likely that the variability could be interpreted in the opposite way: pavements in bad conditions nowadays could have been in good conditions and then associated to comparatively less crashes than bad maintained pavements at the moment of the crashes.

The computed sight distance is another predictor that needs cautious interpretations. For each intersection, the least critical sight distance was determined as the least sight distance among all distances computed on each intersecting leg, considering surrounding obstacles. The critical sight distance was included in the selected models for three and four-legged intersections and in signalized intersections with similar positive coefficients. This can be explained by drivers being less cautious e.g., tending to speeding (e.g., *17* for the rural case) when having more available sight distance at intersections, especially if signalized. This seems not counterintuitive. In fact, it is paramount that sight distance is one of the most important pre-requisite to be guaranteed for a correct road design. However, in specific conditions, such as signalized intersections, a longer sight distance may lead to a false sense of increased safety. This may lead in turn to possible aberrant behaviors, such as running the red light, which is easier when the sight view is unobstructed. Manual explorations of crash dynamics at signalized intersections in the dataset seem to confirm this possibility.

The last variable which is worth of being mentioned is the presence of cycle paths (for segments) and crossings (for intersections). It should be mentioned that all the segments/intersections in the dataset were not originaly designed with cycle paths/crossings, but were recently implemented on the city network. The presence of bycicle crossing is associated with an increase in the number of crashes for the models for three-legged and unsignalized intersections. This is an indication that intersections with vulnerable users such as cyclists should be improved, e.g., with the implementation of specific traffic calming measures, which are not present in the network studied (except for crossing being colored in red). This implementation should be prioritized at unsignalized intersections. Moreover, the presence of bus stops close to signalized intersections seems to be positive for safety. This could be associated with drivers being forced to slow down for the combined presence of bus stops and intersections.

CONCLUSIONS

Safety performance functions for urban segments and intersections were developed in this study as based on datasets from Bari, Italy. The study was conceived for both research purposes and for being useful for road safety practitioners. The research aims were dedicated to: a) explore possible different subsets of segments and intersections with the aim of detailed crash predictions, by assessing also the related variability of predictors, b) compare results from the study with relevant literature, to find common trends and/or significant differences, c) assess the importance of additional predictors often not included in crash modelling.

The optimal subsets found in this study were: one-way and two-way segments for the dataset of homogeneous segments; three-legged, four-legged, unsignalized, and signalized intersections for the dataset of intersections. Hence, besides of overall models for the whole segments and intersections, six detailed models were estimated for the defined subsets. The division into three-legged/four-legged intersections is the most effective, with respect to the signalized/unsignalized intersections. Whereas, significant differences were not highlighted for the number of lanes on one-way and two-way segments. In terms of predictors, intersection models share some common trends with previous similar studies (*2, 12*) and some notable differences (*12*). However, the segment model is largely different than the relevant reference study analysed (*1*).

Some additional predictors often not included in prediction models were found as statistically significant in the present study. In particular, the effect of pavements/marking maintenance, critical sight distance at intersections, vertical signs on driveways/minor roads and cycle path crossings was discussed in detail. While their interpretation has required some explanations, their influence on crash predictions was clearly demonstrated. Hence, it is adviced to use this information in further studies for urban safety predictions.

The results shown in this study can be used for safety predictions in the same area in which data were collected, where the development and use of SPFs is surely needed. However, it is likely that they could be generalized in other Italian/European cities with similar configurations, as based for example on some similarities shown with the Portuguese models (*2*). However, their transferability may provide a challenge. Besides their practical use, it provides new insights to overcome the problems and consequences of dividing urban intersections and segments into different possible subsets and to increase the number of candidate predictors for crash frequencies. Clearly, this study is based on a limited number of segments and intersections with small a number of crashes some subsets, which may negatively influence the crash predictions (*18*). Further data collection is currently in progress during the research project, which could help to enlarge datasets and it will also be useful to validate the models and the selected variables. Moreover, roundabouts were not considered since only a few roundabouts were present in the city during the observation period. The explained variability of crash frequencies is someway small for some of the models conducted. Thus, there are several other variables which may be considered for safety predictions. A first attempt to enlarge them was conducted here, even if further research is needed.

ACKNOWLEDGMENTS

This City of Bari, main leader of the Pa.S.S.S. research project, funded by the Italian Ministry of Infrastructures and Transport, main scientific partner: Politecnico di Bari (grantholder: Prof. Colonna), is acknowledged. Students of the “Road Safety” course (Politecnico di Bari, chair: Prof. Colonna) are acknowledged for their contribute to the early data collection.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Paolo Intini, Pasquale Colonna, Nicola Berloco, Vittorio Ranieri; data collection: Gabriele Cavalluzzi; analysis and interpretation of results: Paolo Intini, Pasquale Colonna, Gabriele Cavalluzzi, Dominique Lord, Vittorio Ranieri; draft manuscript preparation: Paolo Intini, Nicola Berloco, Dominique Lord. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Greibe, P. Crash prediction models for urban roads. *Accident Analysis & Prevention*, 35(2), 273-285. 2003.

2. Gomes, S. V., Geedipally, S. R., & Lord, D. Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety science*, *50*(9), 1732-1739. 2012.

3. Hauer, E., Council, F. M., & Mohammedshah, Y. Safety models for urban four-lane undivided road segments. *Transportation research record*, *1897*(1), 96-105. 2004.

4. Vogt, A., & Bared, J. Crash models for two-lane rural segments and intersections. *Transportation Research Record*, *1635*(1), 18-29. 1998.

5. Persaud, B., & Dzbik, L. Crash prediction models for freeways. *Transportation Research Record*, *1401*, 55-60. 1992.

6. AASHTO. Highway Safety Manual. 1st Edition. 2010.

7. Geedipally, S.R., Shirazi, M., Lord, D. Exploring the need for region-specific calibration factors. *Transportation Research Record* 2636, 73–79. 2017.

8. Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., & Wang, J. H. Exploring the transferability of safety performance functions. *Accident Analysis & Prevention*, *94*, 143-152. 2016.

9. Sacchi, E., Persaud, B., & Bassani, M. Assessing international transferability of highway safety manual crash prediction algorithm and its components. *Transportation research record*, *2279*(1), 90-98. 2012.

10. Colonna, P., Intini, P., Berloco, N., & Ranieri, V. Integrated American-European protocol for safety interventions on existing two-lane rural roads. *European transport research review*, *10*(1), 5. 2018.

11. Colonna, P., Intini, P., Berloco, N., Fedele, V., Masi, G., & Ranieri, V. An Integrated Design Framework for Safety Interventions on Existing Urban Roads—Development and Case Study Application. *Safety*, *5*(1), 13. 2019.

12. Fancello, G., Soddu, S., & Fadda, P. A crash prediction model for urban road networks. *Journal of Transportation Safety & Security*, *10*(4), 387-405. 2018.

13. Canale, S., Leonardi, S., & Pappalardo, G. The reliability of the urban road network: Crash forecast models. In *Proceedings of the III International Congress SIIV-People, Land, Environment and Transport Infrastructures, Bari, Italy* (pp. 22-24). 2005.

14. Jonsson, T. *Predictive models for crashes on urban links-A focus on vulnerable road users*. Bulletin/Lund Institute of Technology, Department of Technology and Society, *226*. 2005.

15. Lord, D., & Mannering, F. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation research part A: policy and practice*, *44*(5), 291-305. 2010.

16. Venables, W. N., & Ripley, B. D. *Modern applied statistics with S*. Springer, New York, NY. 2002.

17. Colonna, P., Intini, P., Berloco, N., & Ranieri, V. The influence of memory on driving behavior: How route familiarity is related to speed choice. An on-road study. *Safety science*, *82*, 456-468. 2016.

18. Lord, D. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention*, *38*(4), 751-766. 2006.