THE INFLUENCE OF TRAFFIC, GEOMETRIC AND CONTEXT VARIABLES ON URBAN CRASH TYPES: A GROUPED RANDOM PARAMETER MULTINOMIAL LOGIT APPROACH

by Paolo Intini[1,*], Nicola Berloco[1], Achille Fonzone[2], Grigorios Fountas[2], Vittorio Ranieri[1]

[1]Department of Civil, Environmental, Building, Land Engineering and Chemistry
Polytechnic University of Bari
4 via Orabona, Bari 70126, Italy

[2]Transport Research Institute,
Edinburgh Napier University
10 Colinton Road, Edinburgh EH10 5DT, United Kingdom

*Corresponding author, contacts: paolo.intini@poliba.it, +39 0805963390

ABSTRACT

Numerous road safety studies have been dedicated to the estimation of crash frequency and injury severity models. However, previous research has shown that different factors may influence the occurrence of crashes of different types. In this study, a dataset including information from crashes occurred at segments and intersections of urban roads in Bari, Italy was used to estimate the likelihood of occurrence of various crash types. The crash types considered are: single-vehicle, angle, rear-end and sideswipe. Models were estimated through a mixed logit structure considering various crash types as outcomes of the dependent variable and several traffic, geometric and context-related factors as explanatory variables (both site- and crash-specific). To account for systematic, unobserved variations among the crashes occurred on the same segment or intersection, the grouped random parameters approach was employed. The latter allows the estimation of segment- or intersection-specific parameters for the variables resulting in random parameters. This approach allows assessing the variability of results across the observations for individual segments/intersections.

Segment type and the presence of bus lanes were included as explanatory variables in the model of crash types for segments. Traffic volume per entering lane, total entering lanes, total number of zebra crossings and the balance between major and minor traffic volumes at intersections were included as explanatory variables in the model of crash types for intersections. Area type was included in both segment and intersection models. The typical traffic at the moment of the crash (from on-line traffic prediction tools) and the period of the day were associated with different crash type likelihoods for both segments and intersections. Significant variations in the effect of several predictors across different segments or intersections were identified. The applicability of the study framework is demonstrated, in terms of identifying roadway sites with anomalous tendencies or high-risk sites with respect to specific crash types.

Keywords: road safety; crash types; grouped random parameters; multinomial logit; urban segments; urban intersections.

1. Introduction

Urban road crashes result in about 15,000 deaths per year in the European Union only (EU-28: 1999–2014 Eurostat data). A recent study (Bauer et al., 2016) has pointed out that urban road fatalities are decreasing over time in the EU, but their percentage among all crashes is nearly stable (actually, it is slightly increasing). Moreover, in some South/Eastern European countries and Portugal (see Bauer et al., 2016) fatalities caused by urban crashes account for more than half of the total fatalities. In the United States, the number of urban fatalities is even increasing, on average, considering a 10-year trend until 2017, and they have exceeded the number of rural fatalities over the recent years (NHTSA, 2019). Since the crash involvement rate of vulnerable road users is notable in urban environments (especially in serious-injury crashes, see Aarts et al., 2016), the need for safer cities (in particular for vulnerable road users) requires thorough understanding of the generation mechanism of severe urban crashes.

There is a considerable amount of research in the field of crash frequency modelling for urban road segments and intersections (Sayed and Rodriguez, 1999; Lord and Persaud, 2000; Persaud et al.; 2002; Harwood et al., 2007). However, as highlighted in Colonna et al. (2019a), most of them concern urban roads in the U.S., which may be significantly different than European urban environments. Transferability issues of models from the U.S. to European contexts (and even within the same country) were already raised indeed (Sacchi et al., 2012; Colonna et al., 2018). Some instances of European urban crash prediction models are anyway present in literature (e.g. Greibe, 2003; Gomes et al., 2012; Intini et al., 2019a). As well as crash frequency modelling, there is a considerable amount of research concerning injury severity modelling with different techniques (see e.g., Kockelman and Kweon, 2002; Abdel-Aty, 2003; Malyshkina and Mannering, 2009; Savolainen et al., 2011; Yasmin and Eluru, 2013; Russo et al., 2014; Yasmin et al., 2014; Fountas and Anastasopoulos, 2017; Fountas et al., 2018a, Behnood and Mannering, 2019). However, also in the case of severity models, most studies were conducted with data from the U.S. and by considering the rural or mixed urban/rural environment.

Besides modelling crash frequency and crash severity, previous research (Kim et al., 2006, 2007; Jonsson et al., 2007, 2009) has shown the importance of differentiating crashes into crash types, in order to highlight variations in the influence of traditional predictors. However, the latter aspect is often overlooked in crash frequency and crash severity analyses, especially in urban environments. For instance, all the above cited studies (Kim et al., 2006, 2007; Jonsson et al., 2007, 2009) refer to rural intersections. The importance of differentiating crashes considering crash types and studying differences between influential predictors is also crucial for identifying specific countermeasures, which can be effective for a given crash type (see e.g., Retting et al., 1995). In fact, some countermeasures can generally improve safety performances, e.g., those aimed at reducing speeds leading, in turn, to crash reduction (Aarts and Van Schagen, 2006; Elvik, 2013). However, some other are specifically targeted at addressing some specific crash types. For example, if there is a significant amount of angle crashes at signalized intersections, then traffic light systems could be improved (e.g., by implementing dedicated turn signals, depending on the prevailing traffic flow and the intersection-specific crash patterns). This evidence could not emerge from a traditional crash frequency model or an injury severity analysis.

Hence, this study is focused on the analysis of the predictors of specific urban road crash types. Using a dataset of urban crashes and related site-specific and crash-specific explanatory variables, the probability of a crash of a given type to occur (conditional on a crash having occurred and recorded through a crash report) is modelled. This problem is typically addressed through a multinomial logit structure, in case of non-binary crash outcomes. Multinomial logit structures were extensively used in previous research concerning injury severity analysis (see e.g., Shankar and Mannering, 1996; Tay et al., 2011; Celik and Oktay, 2014), in their standard formulation or with some modifications (e.g., Savolainen and Mannering, 2007; Chen et al., 2015; Wali et al., 2018; Alnawmasi and Mannering, 2019). In some instances, they were also used for predicting different crash type outcomes (Geedipally et al., 2010; Bham et al., 2011; Chen et al., 2016), such as in the present work.

In predictions made through multinomial logit structures, the observational unit is the individual crash. However, multiple crashes can occur on the same segment or intersection. A mixed logit model structure was implemented to capture unobserved heterogeneity, i.e. the effect of the influential factors that are not apparent to the analyst (Mannering et al., 2016). Treating the crash observations individually regardless of the roadway segment or intersection where they crashes occurred could lead to biased predictors as commonly shared variations across crashes occurred on the same segment or intersection cannot be effectively captured (Mannering et al., 2016; Sarwar et al., 2017; Fountas et al., 2018b; Cai et al., 2018). In this study, to address the aforementioned limitation, the model parameters are allowed to vary across groups of segment- or intersection-specific crashes through the estimation of grouped random parameters. Such an approach, used in previous research (Sarwar et al., 2017; Cai et al., 2018, Eker et al., 2019; Heydari et al., 2019, Pantangi et al., 2019), also paves the way for site-specific evaluation of crash risk considering various crash types. Mixed logit models have been consistently applied in accident research, with some individual differences between studies, for injury severity analyses (Milton et al., 2008; Kim et al., 2013; Wu et al., 2014; see Savolainen et al., 2011 for an early review). However, to the authors' knowledge, no previous study has applied the grouped random parameter multinomial logit structure for predicting crash types. As previously discussed, highlighting the specific influence of the considered predictor at the segment/intersection-level may reveal local patterns, which is useful for practical purposes (i.e. selecting specific countermeasures).

The study answers the following main research questions:

- What are the main geometric and traffic-related predictors of crash types on urban segments and intersections?
- Is it possible to associate crash-specific variables (i.e. context variables, not directly related to the geometry of segments and intersections) to different urban crash types?
- Does the influence of predictors on crash types vary considerably across segments or intersections?

Research questions are addressed by analysing a dataset from an Italian city. Considering the aforementioned gaps in previous research, this study, which is exploratory in its nature, expand the existing knowledge in several ways: a) conducting safety analysis disaggregated for different crash types, b) deepening knowledge related to urban road safety predictions, c) highlighting results from the application of a grouped random parameter multinomial logit structure to crash type prediction, d) using a dataset from an European city, considering the impact of urban spatial setting on traffic safety.

The remainder of the paper is structured as follows. Methods used for data analysis are described in detail in the next section. Then the modelling results are presented and discussed, in light of previous relevant research. The applicability of the results is shown in practice, by highlighting specific high-risk sites based on the modelling results. Finally, the main conclusions from the study are drawn.

2. Methods

The methods used in this article are described as follows, starting with the crash dataset and the predictors that were used for the statistical analysis of crash types. Next, the statistical methods used for model estimation are presented in detail.

2.1 Database

The study is part of a larger National research project ("Scientific Park for Road Safety", funded by the Italian Ministry of Transport and Infrastructures, leading agency: Municipality of Bari, Italy). In this project, evidence from local urban road safety studies is used to infer possible policies and strategies, which may help reduce urban crashes at a higher level (e.g., at a national level). In the context of this research project, data about crashes occurred on the road network of the Municipality of Bari between 2012 and 2016 were collected and put together with some possible influential variables, which may be related to crashes. The City of Bari is a medium-sized Southern Italian city, with a population of about 320,000 inhabitants, and an area of about 120 km^2.

Crash data were provided by ASSET (http://asset.regione.puglia.it/), the local agency that manages these data in collaboration with the National Institute of Statistics (ISTAT). In addition to publicly available crash data, the exact localisation of the crash (GPS position) is included in the dataset provided. Note that the crash dataset provided, according to the European state-of-practice, includes only fatal+injury crashes, which are locally collected and standardized by the National Institute of Statistics (ISTAT). The crash dataset includes information about the day, hour, crash type, the involved vehicles and users, the contributory factors and the boundary conditions (i.e., weather, pavement, etc.). Other information was manually matched with crash data instead, such as road geometric data and traffic volumes (more details are provided in: Intini et al., 2019b; Colonna et al., 2019b).

Based on localisation, crash data were assigned to the road segments or intersections. In cases where inaccuracies in the data localisation did not allow to identify the crash site precisely, the records were removed from the initial dataset. Give-way/stop lines and zebra crossings (included in the intersection area if close to the intersections) were initially used as preliminary thresholds for intersection-related crashes. However, given the high probability of misclassification of crashes (into intersection- or segment-related crashes) when the classification is based on fixed thresholds (e.g., distance from the intersection centre or stop lines/crossings position), crash locations, types, circumstances and related features were manually explored, to distinguish the intersection-related crashes from the segment-related crashes. This further level of preliminary analysis was necessary given that this study is focused on crash types, separately assessed for segments and intersections. Moreover, segments were divided into homogeneous sections on the basis of their internal geometric characteristics (e.g., a different number of lanes, or the presence of medians). In other words, if notable macro-differences were identified among different sections of the same segment located between two major intersections (excluding driveways and intersections with minor roads), that segment was split into two or more homogeneous sections (AASHTO, 2010). For this reason, the word “segment” is henceforth referred to as homogeneous sections. Descriptive statistics about crash data are reported as follows, differentiated for segments and intersections of the urban road network.

The study is focused on crash types, and then information about crash types were retrieved from the database. The most disaggregate classes found for crash types are: run-off-road, fixed object, pedestrian hit, fallen from vehicle, angle, head-on, sideswipe (not further classified by vehicle directions), rear-end. Since some of these categories were significantly under-represented in the sample (e.g., the fallen from vehicle crash: only 2 crashes), then crash types were grouped into broader categories. Run-off-road, fixed-object, pedestrian hit and fallen from vehicle crashes were grouped into a “single-vehicle” crash type, given that only one vehicle was involved. Moreover, head-on crashes account for only about 3% of the total sample (29 out of 1036). However, to avoid grouping head-on crashes with other multi-vehicle crash types with significantly different mechanisms, head-on crashes were discharged from the dataset. In the final dataset used for model estimates, there are on average 3.20 fatal+injury crashes per segment (st.dev.: 3.27) and 4.96 fatal+injury crashes per intersection (st.dev.: 4.70).

As far as the site-specific explanatory variables are concerned, segment and intersection types include different combinations of one-way/two-way, single/multilane, undivided/divided segments and signalized/unsignalized, three/four-legged intersections. In this case too, classes of segments and intersections were appropriately formed in order to avoid having classes with very few elements (such as three-legged signalized intersection that comprise only 4 % of all signalized intersections). Average annual daily traffic per lane was used as a measure of traffic exposure. In case of intersections, it should be interpreted as number of vehicles per day per lane entering into the intersection (scaled down by using the unit of measurement: hundreds of vehicles per day per lane for modelling purposes). The ratio between the traffic volume on the major road and the traffic volume on the minor road was computed to capture the balance between the two volumes; the latter has been previously found to be associated with safety issues at intersections (Gomes et al., 2012; Intini et al., 2019b). Other site-specific variables included in the dataset were: segment length, total entering lanes in the intersection, number of zebra crossings (at both segments and intersections), presence of bike paths and bus lanes (on segments), area type, presence of nearby public attractors (i.e., schools; hospitals; governmental buildings; etc.). A continuous measure representing the number of entering lanes was preferred against an indicator variable such as e.g., more or less than four entering lanes, because the latter classification was deemed to assume a higher degree of arbitrariness in the threshold lanes with respect to the continuous variation. However, the authors are not interested here in specifically assessing the effects of each one entering lane increase, but the number of entering lanes was rather used in this study as a proxy measure for the complexity of the intersection. In fact, it is assumed that the complexity can have an influence on different crash type outcomes.

Area type was defined with regard to different city areas, as shown in Fig. 1. The speed limit was consistently equal to 50 km/h for all the sites during the observation period. However, the configuration of the segments and intersections is largely different between the city centre (typically consisting of short segments with several major intersections with low spacing between them) and the rural-to-urban transition areas (typically consisting of long segments with intersections spaced with a notable distance), while neighbourhoods of the city centre are in an intermediate condition. This may significantly affect speed and driving behaviour (Silvano and Bang, 2015; Colonna et al., 2019a), with city centre areas reflecting operating speeds significantly lower than 50 km/h and transition areas reflecting operating speeds significantly higher than 50 km/h. To capture this difference, the area type variable was introduced in the analysis. Segments in sparsely populated areas, which lead to the main beltway connecting to the rural network were assigned to the "transition area" category as well as the intersections lying on them. Moreover, the transition area variable is also used as a surrogate measure of parking, since on most of the sample sites included in this area there is no on-street parking, contrary to the roads belonging to the other area types (city centre and neighbourhoods).

Crash-specific explanatory variables were obtained from the crash dataset. They include basic information such as crash date and hour and pavement conditions at the moment of the crash. Based on this information, the following variables were defined: season, type of day (weekday or weekday/holidays), period of the day (6 a.m.-6 p.m. or 6 p.m.-6 a.m., henceforth referred to as, namely, "day" or "night"), pavement conditions (dry or wet/slippery/icy). Moreover, a qualitative, crash-specific measure of the traffic volume that was present at the moment of the crash was inferred from the online Google Maps® tool for typical traffic at given hours and given days of the week, based on a colour scale (ranging from green labelled as "fast", to dark red: "slow"). Hence, in this study, three classes were defined aggregating information inferred from the colour scale: no delays expected (green colour), some delays expected (orange colour), delayed/congested traffic (red/dark red colours, colours grouped together since there are very few situations in which the dark red colour is observable on the inquired road network). It should be noted that the measure is highly qualitative, since no numerical thresholds were considered and it is based on visual exploration of on-line sources. However, it was deemed as an interesting potential measure for capturing real-time traffic conditions, which are otherwise very hard to obtain (while they are generally useful for safety modelling, see Christoforou et al., 2011; Shi and Abdel-Aty, 2015).

Table 1. Descriptive statistics of crash data and related information collected for the sample of urban road segments and intersections.

Variables	Segments (n=119)		Intersections (n=129)	
	Mean (S.D.)[1]/ Count (%)[1]	Min.-Max.	Mean (S.D.)[1]/ Count (%)[1]	Min.-Max.
General frequency variables				
Fatal+injury crashes	379	-	628	-
Fatal+injury crashes/site	3.19 (3.22)	1-18	4.87 (4.67)	1-29
Differentiated by crash type				
Single vehicle crashes/site	1.04 (1.50)	0-11	0.95 (1.21)	0-7
Angle crashes/site	0.71 (1.22)	0-8	2.37 (2.70)	0-13
Rear-end crashes/site	0.84 (1.40)	0-10	0.75 (1.34)	0-8
Sideswipe crashes/site	0.60 (0.87)	0-5	0.85 (1.35)	0-7
Dependent variable: crash type				
Crash type: *Single-vehicle*	124 (0.33)	-	119 (0.19)	-
Crash type: *Angle*	84 (0.22)	-	323 (0.51)	-
Crash type: *Rear-end*	100 (0.26)	-	83 (0.13)	-
Crash type: *Sideswipe*	71 (0.19)	-	103 (0.16)	-
Explanatory variables: site-specific				
Segment type: *One-lane*	50 (0.13)	-	-	-
Segment type: *Undivided 1-way 2+ lanes*	42 (0.11)	-	-	-
Segment type: *Undivided 2-way 2-lanes*	115 (0.31)	-	-	-
Segment type: *Undivided 2-way 4-lanes*	90 (0.24)	-	-	-
Segment type: *Divided 2-way*	82 (0.22)	-	-	-
Intersection type: *Unsignalized 3 legs*	-	-	118 (0.19)	-
Intersection type: *Unsignalized 4 legs*	-	-	141 (0.22)	-
Intersection type: *Signalized*	-	-	369 (0.59)	-
Segment length (m)	194.4 (169.4)	34-862	-	-
Average traffic per lane [vehicles/day]	4410.7 (2200.6)	250-11460	4002.3 (2196.7)	500-15570
% Ratio: minor to major traffic volume	-	-	47.1 (30.2)	0.0-100.0
Total entering lanes	-	-	4.7 (2.6)	1-11
Number of zebra crossings	0.6 (0.8)	0-3	2.9 (1.3)	0-5
Presence of bus lanes: *No*	342 (0.90)	-	-	-
Presence of bus lanes: *Yes*	37 (0.10)	-	-	-
Presence of bike paths: *No*	344 (0.91)	-	-	-
Presence of bike paths: *Yes*	35 (0.09)	-	-	-
Area type: *Neighbourhood*	230 (0.61)	-	401 (0.64)	-
Area type: *City Centre*	100 (0.26)	-	144 (0.23)	-
Area type: *Transition area*	49 (0.13)	-	83 (0.13)	-
Presence of nearby public attractors: *No*	213 (0.56)	-	338 (0.54)	-
Presence of nearby public attractors: *Yes*	166 (0.44)	-	290 (0.46)	-
Explanatory variables: crash-specific				
Season: *Winter*	88 (0.23)	-	172 (0.27)	-
Season: *Spring*	112 (0.30)	-	172 (0.27)	-
Season: *Summer*	98 (0.26)	-	146 (0.23)	-
Season: *Autumn*	81 (0.21)	-	138 (0.22)	-
Type of day: *Weekday*	304 (0.80)	-	468 (0.75)	-
Type of day: *Weekend/public holiday*	75 (0.20)	-	160 (0.25)	-
Period of the day: *Day (6 a.m.-6 p.m.)*	272 (0.72)	-	392 (0.62)	-
Period of the day: *Night (6 p.m.-6 a.m.)*	107 (0.28)	-	236 (0.38)	-
Typical traffic at crash: *No delays*	107 (0.28)	-	134 (0.21)	-
Typical traffic at crash: *Some delays expected*	214 (0.56)	-	336 (0.54)	-
Typical traffic at crash: *Delayed*	21 (0.06)	-	47 (0.07)	-
Typical traffic at crash: *No available data*	37 (0.10)	-	111 (0.18)	-
Pavement conditions: *Dry*	336 (0.89)	-	544 (0.87)	-
Pavement conditions: *Other*	43 (0.11)	-	84 (0.13)	-

[1]Depending on the variable being numerical or categorical, namely means (with standard deviations S.D. in parenthesis) or counts (with percentages among the total % in parenthesis) are presented.

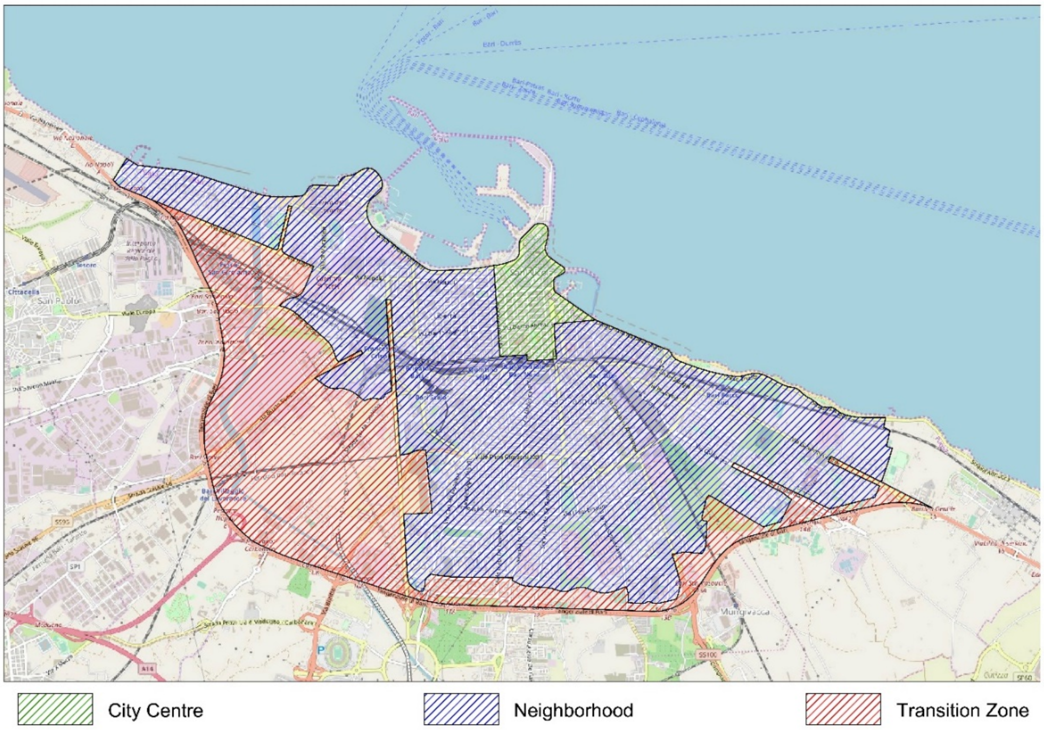


Figure 1. Considered area types in the city of Bari, Italy (source image from OpenStreetMap)

2.2 Statistical methods

In this study, a multinomial logit structure was used to predict the likelihood of different crash types (with four possible outcomes: single-vehicle, angle, rear-end, sideswipe). The most disaggregate observational unit used for modelling is the individual crash in the dataset. Site-specific and crash-specific explanatory variables are used to predict the likelihood of different crash types. Note that, based on the data availability and sample size, the crash type outcome was chosen as dependent variable, rather than crash frequency by crash type (with road sites as observational units, see Mothafer et al., 2016; Bhowmik et al., 2019) or proportion of crashes (applied at a macro-level by Lee et al., 2018).

Two separate models were developed for the segment and intersection datasets. Instead of the standard multinomial logit approach (previously used for similar purposes by Geedipally et al., 2010; Bham et al., 2011; Chen et al., 2016), a mixed (random-parameter) logit structure was preferred. In fact, this approach enables the model parameters to vary across the different units (Washington et al., 2020; Mannering et al., 2016). In this specific case, the parameters are allowed to vary across the segments or intersection. As such, rather than having a single parameter estimate for each individual crash, the parameters were grouped for each set of crashes corresponding to each individual segment or intersection. In this way, it may be possible to capture some specific unobserved characteristics (Mannering et al., 2016; Fountas et al., 2018b) of segments and intersections, which could be unfeasible with fixed parameter estimates (i.e., the same coefficient for all segments and intersections).

Let assume the systematic component $V_{ct,c}$ of the likelihood of a given crash type *t* for a crash observation *c* as a linear combination of a given set of predictors, in which some of the coefficients may be fixed and some other may be site-specific (segment or intersection-specific):

$$V_{t,c} = \boldsymbol{\beta_i}\, \boldsymbol{X_{t,c}} + \boldsymbol{\beta_{i,s}}\, \boldsymbol{Z_{t,c}} \tag{1}$$

Where:

$\boldsymbol{\beta_i}, \boldsymbol{\beta_{i,s}}$ = vectors of coefficient estimates associated to the *i*-th predictor which are, namely, fixed and specific to the given site *s*;

$\boldsymbol{X_{t,c}}, \boldsymbol{Z_{t,c}}$ = vectors of predictors of a given crash type *t* likelihood associated to, namely, fixed and site-specific coefficient estimates.

In this case, the probability of observing a crash type outcome *t* estimated through a mixed logit model structure can be defined as follows (adapted from Milton et al., 2008; Washington et al., 2020):

$$P_c(t) = \int_{\boldsymbol{X}} \frac{\exp(\boldsymbol{\beta_t X_{t,c}})}{\sum_T \exp(\boldsymbol{\beta_t X_{t,c}})} f(\boldsymbol{\beta}|\boldsymbol{\theta})\, d\boldsymbol{\beta} \tag{2}$$

Where:
$P_c(t)$ = probability of observing the crash type outcome *t* (among the set of crash type outcomes T) for the crash unit *c*;
$\boldsymbol{\beta_t}$ = vector of estimated parameters for the different crash types *t*;
$\boldsymbol{X_{t,c}}$ = vector of explanatory variables for different crash types *t,* for the crash unit *c*;
$f(\boldsymbol{\beta}|\boldsymbol{\theta})$ = probability density function assumed for $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ is the vector of parameters of the function.

In this study, a grouped random parameter approach (Sarwar et al., 2017; Cai et al., 2018) was used: individual parameters β are estimated for each group of crashes occurred at each segment or intersection. Moreover, a normal distribution was assumed for the density function $f(\boldsymbol{\beta}|\boldsymbol{\theta})$, in line with results from previous research (e.g., Milton et al., 2008; Moore et al., 2011). Note that several of the explanatory variables are categorical (see Table 1). Thus, in this case, binary dummy variables were generated (1 - presence of the given attribute, 0 - absence of the given attribute, e.g., for winter season: 1 - winter, 0 - other seasons).

The *mixlogit* command implemented in the STATA® software (based on Hole, 2007) was used for estimating the mixed logit models. The underlying software algorithm, based on a mathematical transformation from the standard mixed logit structure, estimates the logarithm of the odds of a given outcome with respect to a reference outcome (StataCorp, 2015) in the set, as follows:

$$\ln\left[\frac{P_c(t)}{P_c(t_0)}\right] = \beta_{0,s} + \sum_{i=1}^{\boldsymbol{X_t}} \beta_i X_{t,c} + \sum_{i=1}^{\boldsymbol{Z_t}} \beta_{i,s} Z_{t,c} \tag{3}$$

Where:
$P_c(t_0)$ = probability of observing the reference crash type t_0 (among the set T) for the crash unit *c*;
all other terms were previously defined for Equations 1 and 2. Note that the estimate $\beta_{0,s}$ for the intercept may eventually be site-specific as well, or fixed (β_0).

This approach was previously applied for similar purposes (i.e., crash types as outcomes) in a standard multinomial logit structure (Geedipally et al., 2010; Bham et al., 2011; Chen et al., 2016). Based on Eq. 3, and considering that the sum of the observed probabilities of all outcomes should be equal to 1, the probability of observing each crash type outcome *t* can be computed. In this case, using the above explained transformation for the model application leads to estimating three functions, by selecting the single-vehicle crash type as a reference.

According to literature, the mixed logit model was developed using a maximum likelihood estimation approach coupled with the Halton draws sampling technique (Halton, 1960). The models presented in this study were generated using 1000 Halton draws, in line with numbers effectively used in previous research (Milton et al., 2008; Moore et al., 2011; Kim et al., 2013; Wu et al., 2014). The model selection process was conducted by trying to simultaneously include only predictors for which the estimated coefficients are statistically significant at the 10 % level, given the small dataset and the exploratory nature of this study. Moreover, the Akaike Information Criterion (AIC) was also computed and evaluated to compare different models.

To assess the impact of each predictor included in the model functions on the outcome probabilities, elasticities were computed. Depending on the results from the model, different predictors can be included in one or more functions related to different crash types. For this reason, both direct and cross point elasticities were computed for each crash unit, starting from the initial dataset. For a one percent change in the predictor, the point elasticities represent the percentage difference in the outcome probability (Washington et al., 2020), defined as follows:

$$E_{X_{t=i,c}}^{P(t=i)} = \frac{\Delta P(t=i)}{P(t=i)} * 100\ (\%) \tag{4}$$

$$E_{X_{t=i,c}}^{P(t=j)} = \frac{\Delta P(t=j)}{P(t=j)} * 100\ (\%) \tag{5}$$

Where:

$E_{X_{t=i,c}}^{P(t=i)}$= direct elasticity, percent change in the probability $P(t = i)$ of observing the crash type *i,* for a one percent increase in the predictor $X_{t=i,c}$, included in the function associated to the crash type *i.*

$E_{X_{t=i,c}}^{P(t=j)}$= cross elasticity, percent change in the probability $P(t = j)$ of observing the crash type *j,* for a one percent increase in the predictor $X_{t=i,c}$, included in the function associated to the crash type *i.*

Elasticities were computed by applying the model functions and the estimated set of individual parameters for each segment and intersection, in case of random parameters; and the mean estimate in case of fixed parameters. In case of binary predictors, pseudo-elasticities were computed (Washington et al., 2020). The formulation of pseudo-elasticities is similar to the previous equations; instead of the effect of a one percent change, the effect of a change in the dummy variable from 0 to 1 is estimated for all the observations. Once elasticities and pseudo-elasticities are estimated for each crash unit, average elasticities are computed among the observations, to represent an overall effect.

3. Results

The results for the separate sub-sets of segment- and intersection-related crashes are reported in this section and discussed in the following one.

3.1 Model for segment crashes

The predictors and the related estimated coefficients associated to different crash types likelihood on segments (with respect to single-vehicle crashes) are presented in Table 2.

Table 2. Estimated model for segment crashes

Explanatory variables	**Coefficient (st. dev.)***	**St. error^**	***p*-value^**	**Lower value 95 % C.I.^**	**Upper value 95 % C.I.^**
Reference crash type: Single vehicle crashes					
Crash type: Angle					
Undivided 2-way 4-lane segment	1.048	0.305	0.001	0.450	1.645
Area type – City centre	-1.296	0.363	<0.001	-2.008	-0.584
Typical traffic – Some delays expected	-0.377	0.197	0.055	-0.763	0.008
Crash type: Rear-end					
Area type – Transition area	1.910	0.321	<0.001	1.281	2.538
Night (6 p.m.-6 a.m.)	-0.750	0.260	0.004	-1.260	-0.239
Crash type: Sideswipe					
Presence of bus lanes	-1.545	0.676	0.022	-2.869	-0.221
Night (6 p.m.-6 a.m.)	-2.200 (2.741)	1.272 (1.475)	0.084 (0.063)	-4.692(-0.149)	0.246 (5.632)
Goodness-of-fit					
AIC = 983.44, *LL(**β**)* = -483.72					
Wald test: $\chi^2(7) = 59.88$, $p < 0.001$.					
Likelihood Ratio Test (comparison with the correspondent fixed parameters model): $\chi^2(1) = 7.01$, $p = 0.008$.					
In-sample predictions					
Crash type outcome for each crash in the dataset, correct choices[+]: 276 (73%), incorrect choices: 103 (27%)					
Most frequent crash type for each segment (aggregated choices), correct[+]: 100 (84%), incorrect: 19 (16%)					

*Values in parenthesis are the estimated standard deviations of coefficients in case of estimated random parameters.
^Values in parenthesis are computed for the estimated standard deviations of coefficients in case of random parameters.
+A correct choice was assumed if the predicted outcome matched the observed outcome (the most frequent outcome, even paired with other equiprobable outcomes).

Predictors included in the model are: the segment type (undivided 2-way 4-lane segments in case of angle crashes), the area type (city centre in case of angle crashes, transition areas in case of rear-end crashes), the typical traffic (some delays expected in case of angle crashes), the day period (in case of both rear-end and sideswipe crashes). Traffic volume and segment length were not included as predictors in the model, due to the lack of statistically significant estimates, as well as several other segment-specific and crash-specific variables.

The coefficient for the period of the day (night: 6 p.m.-6 a.m.) in the function of sideswipe crashes likelihood (with respect to single vehicle crashes) was estimated as a random parameter across the segments. This means that, given the approach selected, a specific coefficient estimate is calculated for each segment. The grouped random parameter approach leads to a statistically significant improvement with respect to the correspondent fixed parameters model (i.e., considering a fixed parameter for the period-of-the-day variable in the function of sideswipe crashes), as based on the Likelihood Ratio Test (LRT - see Table 2); the latter reveals an overall significance for the estimated standard deviation (Hole, 2007). Moreover, the Wald test confirms that the selected predictors included in the model significantly improve the fit.

Based on the estimates presented in Table 2, elasticities are computed in Table 3. Given that all the predictors included in the segment model are indicators, then pseudo-elasticities are computed.

Table 3. Pseudo-elasticities computed for all crash type outcomes T (Single-Vehicle: SV, Angle: AN, Rear-end: RE, Sideswipe: SS) – segment model

Explanatory variables	Percentage change in Probability of each crash type (%)			
	Single Vehicle (SV)	Angle (AN)	Rear-end (RE)	Sideswipe (SS)
Undivided 2-way 4-lane segment	-24.4*	115.7	-24.4*	-24.4*
Presence of bus lanes	19.5*	19.5*	19.5*	-66.8
Area type – City centre	27.6*	-65.1	27.6*	27.6*
Area type – Transition area	-57.5*	-57.5*	186.7	-57.5*
Typical traffic – Some delays expected	9.0*	-17.6	9.0*	9.0*
Night (6 p.m.-6 a.m.)	51.3*	51.3*	-28.5	-60.1

*Cross elasticities. If a given variable is included in only some functions related to specific crash types, then elasticities are computed for these crash types only. Since the sum of choice probabilities should be equal to 1, the probabilities related to the other crash types for which the given variable is not included in the respective functions will decrease/increase of the same quantity accordingly, given the definition of cross-elasticity itself.

Based on the computed pseudo-elasticities, the effects of several variables are further highlighted. There is a significant increase (+116%) in the probability of observing angle crashes on undivided 2-way 4-lane segments. There is also a notable increase (+187%) in the probability of observing rear-end crashes in transition areas. The presence of bus lanes on segments is associated with a decrease (-67%) in the probability of sideswipe crashes, while there is a notable decrease (-65%) in the probability of angle crashes in the city centre. The night period leads to a decrease in the probability of observing sideswipe (-60%) and rear-end (-29%) crashes, while an increase in both probabilities of single vehicle and angle crashes. Minor effects can be noted for the influence of typical traffic with some delays expected on angle crash likelihood (-18%).

3.2 Model for intersection crashes

The predictors and the estimated coefficients associated to the likelihood of different crash types on intersections (with respect to single-vehicle crashes) are presented in Table 4.

Table 4. Estimated model for intersection crashes

Explanatory variables	Coefficient (st. dev.)*	St. error^	*p*-value^	Lower value 95 % C.I.^	Upper value 95 % C.I.^
Reference crash type: Single vehicle crashes					
Crash type: Angle					
Traffic volume per entering lane	0.011 (-0.011)	0.004 (0.004)	0.011 (0.013)	0.002 (-0.019)	0.019 (-0.002)
% Ratio minor-to-major traffic	0.010	0.003	<0.001	0.004	0.015
Typical traffic – Some delays expected	-0.497	0.195	0.011	-0.879	-0.116
Typical traffic – Delayed	-1.356	0.384	<0.001	-2.109	-0.604
Area type – Transition area	2.723	0.745	<0.001	1.262	4.183
Night (6 p.m.-6 a.m.)	0.583	0.197	0.003	0.197	0.970
Crash type: Rear-end					
% Ratio minor-to-major traffic	-0.013	0.003	<0.001	-0.020	-0.007
Area type – Transition area	2.999	0.750	<0.001	1.530	4.468
Night (6 p.m.-6 a.m.)	-1.104 (1.158)	0.597 (0.575)	0.065 (0.044)	-2.274 (-2.285)	0.067 (-0.031)
Crash type: Sideswipe					
Traffic volume per entering lane	-0.008	0.004	0.070	-0.017	0.001
Total entering lanes	0.133	0.044	0.002	0.047	0.219
Area type – Transition area	1.590	0.786	0.043	0.050	3.130
Total zebra crossings	-0.197	0.087	0.024	-0.368	-0.026
Goodness-of-fit					
AIC = 1451.86, *LL(β)* = -710.93					
Wald test: $\chi^2(13) = 174.71$, $p < 0.001$.					
Likelihood Ratio Test (comparison with the correspondent fixed parameters model): $\chi^2(2) = 7.14$, $p = 0.028$.					
In-sample prediction					
Crash type outcome for each crash in the dataset, correct+ choices: 338 (54%), incorrect choices: 290 (46%)					
Most frequent crash type for each segment (aggregated choices), correct+: 94 (73%), incorrect: 35 (27%)					

*Values in parenthesis are the estimated standard deviations of coefficients in case of estimated random parameters.
^Values in parenthesis are computed for the estimated standard deviations of coefficients in case of random parameters.
+A correct choice was assumed if the predicted outcome matched the observed outcome (the most frequent outcome, even paired with other equiprobable outcomes).

Predictors included in the model are: the traffic volume per entering lane (in case of both angle and sideswipe crashes), the ratio of the minor to the major traffic volumes (for both angle and rear-end crashes), the total number of entering lanes (for sideswipe crashes), the total number of zebra crossings (for sideswipe crashes), the typical traffic (both some delays expected and delayed traffic in case of sideswipe crashes), the area type (transition areas for all crash types), the day period (in case of both angle and rear-end crashes). In this case, some intersection-related, traffic and geometric variables are included in the selected model. However, the intersection type (with respect to traffic signals and legs) is not included, while the total number of entering lanes, which reflects the degree of complexity of the intersection, is a predictor of SS crash likelihood (compared to single vehicle crashes).

The coefficients for traffic volume per entering lane (in the angle function) and for day period (in the rear-end function) were estimated as random parameters across the intersections. Given the approach selected, a single coefficient estimate for the two above listed predictors is then obtained for each intersection. The grouped random parameter approach leads to a statistically significant improvement with respect to the correspondent fixed parameters model, as based on the LRT test (see Table 4) which reveals an overall significance for the estimated standard deviations (Hole, 2007). Moreover, the Wald test confirms that the selected predictors included in the model significantly improve the fit.

Based on the estimates presented in Table 4, elasticities are computed in Table 5. In this case, some predictors included in the model are indicator variables and some other predictors are numerical variables. Hence, both elasticities and pseudo-elasticities are computed.

Table 5. Elasticities and pseudo-elasticities computed for all crash type outcomes T (Single-Vehicle: SV, Angle: AN, Rear-end: RE, Sideswipe: SS) – intersection model

Explanatory variables	Percentage change in Probability of each crash type (%)			
	Single vehicle (SV)	Angle (AN)	Rear-end (RE)	Sideswipe (SS)
Elasticities				
Traffic volume per entering lane	-0.2*	0.2	-0.2*	-0.5
% Ratio minor-to-major traffic	-0.2*	0.3	-0.8	-0.2*
Total entering lanes	-0.1*	-0.1*	-0.1*	0.5
Total zebra crossings	0.1*	0.1*	0.1*	-0.5
Pseudo-elasticities				
Area type – Transition area	-90.5*	45.3	91.5	-53.2
Typical traffic – Some delays expected	29.3*	-21.3	29.3*	29.3*
Typical traffic – Delayed	71.5*	-55.8	71.5*	71.5*
Night (6 p.m.-6 a.m.)	-19.0*	45.1	-66.9	-19.0*

*Cross elasticities. If a given variable is included in only some functions related to specific crash types, then elasticities are computed for these crash types only. Since the sum of choice probabilities should be equal to 1, the probabilities related to the other crash types for which the given variable is not included in the respective functions will decrease/increase of the same quantity accordingly, given the definition of cross-elasticity itself,

The effects of variables can be appreciated by considering elasticities and pseudo-elasticities. As far as the numerical variables are concerned, all the relative changes in the outcome probabilities can be considered inelastic (i.e., less than 1% change, see Washington et al., 2020). The most notable effect is the decrease of rear-end crash likelihood in case of consistent traffic volumes across the intersecting legs. The increase in the traffic volume per entering lane is associated with a decrease in the sideswipe crash likelihood and a minor increase in the angle crash likelihood. The sideswipe crash likelihood increases with the total number of entering lanes and slightly decreases with the total number of zebra crossings. Focusing on the indicator variables, there is a notable increase (+92%) in the rear-end crash likelihood for intersections in transition areas, while the single vehicle crash likelihood notably decreases (-91%) as well as the sideswipe crash likelihood, but to a minor extent (-53%). The delayed typical traffic is associated with a decrease in the angle crash likelihood and a notable increase (+72%) in all other crash type likelihoods. The night period leads to a significant decrease (-67%) in the probability of observing rear-end crashes and to an increase in the angle crash likelihood. The effect of a one-unit change of the variable representing typical traffic with some delays expected is minor, resulting in a small decrease in the angle crash likelihood and a correspondent increase in all other crash type likelihoods.

4. Discussion

Herein, the results presented in the previous section are discussed, by following the order of the research questions: a) exploratory analysis of geometric and traffic-related predictors of crash types at urban segments and intersections, b) association of crash-specific variables to urban crash types, c) possible site-specific influential characteristics of given individual segments or intersections.

4.1 Predictors of urban segment and intersection crash types

Several traffic, geometric and context related factors were investigated as potential predictors of different urban crash types likelihood. Among these variables, the presented models include: a) for intersections, the traffic volume per entering lane, the overall number of entering lanes, the total number of zebra crossings and the balance between major and minor traffic volumes; b) for segments, the segment type and the presence of bus lanes; c) for both segments and intersections, the area type context variable. Most of the influential geometric variables are specific to the considered road element (i.e., segments or intersections) and so, their influence is separately discussed for the two road element categories.

For what concerns segments, the undivided 2-way 4-lane segments are associated to an evident increase in the probability of observing an angle crash. This could be attributed to two possible mechanisms. Firstly, speeds may be higher on these urban arterial roads because of the increased road width (as highlighted, for example, by Silvano and Bang, 2015, for free flow speeds). Secondly, vehicles entering from/to driveways/minor intersections should cross more than one lane to turn left (regardless of whether this manoeuvre is allowed, this can occur because they are 2-way multilane roadways not provided with median). The combination of these two factors may explain the higher percentage of angle crashes. The presence of bus lanes is found to be related to a notable decrease in the sideswipe crash likelihood. This can be explained by the lower possibility of lane-changing manoeuvres (which should be considered in detail in urban environments, see Sun and Elefteriadou, 2012) when driving next to lanes dedicated to public transport. This may suggest the use of bus lanes as buffer zones in case of potential sideswipe crashes. Note that the bus lanes in the study area are mostly present on two-lane undivided roads, and some of them are two-way roadways (i.e. with a contraflow bus lane).

For what concerns intersections, the sideswipe crash type likelihood decreases when the traffic per lane entering at the intersection and the total number of zebra crossings increase, while it increases with the number of entering lanes. These results can be explained in parallel. In fact, as the number of entering lanes increases, the possibility of vehicles approaching the intersection on parallel lanes (which may be related to sideswipe crashes, as highlighted by Ackeret et al., 1999, in case of complex turning lane configurations) increases; the latter may increase the probability for lane-changing (e.g., for reaching dedicated turning lanes) and overtaking manoeuvres. However, in cases where the traffic volume per lane increases or in the vicinity of zebra crossings, those manoeuvres can be more difficult to undertake, thus leading to a decrease in the sideswipe crash likelihood. In addition, a decrease in the rear-end crash likelihood is observed in cases where minor traffic volumes are getting closer to the major volumes. This could be explained by drivers reducing speeds and adjusting headways when traffic is balanced among the intersection legs, because of the intrinsic intersection complexity. In fact, it was shown that, as the intersection complexity decreases, inadequate drivers' attention allocation can be suggested, leading to more crashes (Werneke and Vollrath, 2012). Table 5 shows that higher traffic volumes and greater minor-to-major traffic ratios increase the likelihood of angle crashes at intersections. Both identified effects can be explained by the increased number of crossing conflicts, which may generate angle crashes.

Besides of road element-specific geometric variables, there are some variables that were taken into account for both segment and intersection models. Their association with the likelihood of different crash types is shown in Table 6, based on the computed elasticities and pseudo-elasticities in Tables 3 and 5. The influence of traffic per entering lane and total zebra crossings was previously discussed. It is worth to note here that these factors were not found to be influential on the likelihood of different crash types in the segment-based model.

Table 6. Summary of the association of traffic, geometric and context variables to different urban crash type T (Single Vehicle = SV, Angle = AN, Rear-End = RE, Sideswipe = SS) likelihood, common to segments and intersections (S = Segments, I = Intersections)

Common traffic, geometric, and context variables	**Change in Probability of each crash type***							
	Single vehicle (SV)		**Angle (AN)**		**Rear-end (RE)**		**Sideswipe (SS)**	
	S	**I**	**S**	**I**	**S**	**I**	**S**	**I**
Traffic per entering lane^		-		+		-		-
Total zebra crossings^		+		+		+		-
Area type – *City centre*^	+		- -		+		+	
Area type – *Transition area*^	- -	- -	- -	+	+ + +	+ +	- -	- -

**The sign "+" reflects a positive effect (i.e., the specific crash type likelihood is increasing), while the sign "–" reflects a negative effect (i.e., the specific crash type likelihood is decreasing).*

^Numbers of + and - reflect the magnitude of the pseudo-elasticities (+/- for up to ± 50% change, ++/-- for a change included between ± 50% and ± 100%, +++/--- for more than ± 100% change).

The likelihood of different crash types changes if segments and intersections are located in the rural-to-urban transition areas. In both segments and intersections, a notable decrease in the single vehicle and sideswipe crash likelihoods and a notable increase in the rear-end crash likelihood are noted. If the drivers are not guided in the transition from the rural to the urban environment through appropriate design measures (see e.g. Lantieri et al., 2015), they may maintain a typically rural-based driving behaviour (Colonna and Berloco, 2011). In this case, the sub-urban characteristics of these road segments and intersections may allow drivers to maintain high speeds (see Liu, 2007 in case of approaching intersections) but also provide the ground for aggressive driving behaviour, possibly due to the presence of mind wandering and distraction (for further details, see also Fountas et al., 2019). Such behavioural trends are typically observed in low-demand roadway environments (Lin et al., 2016), such as e.g., low traffic rural highways. This may explain the increase in the rear-end crash likelihood. On the other hand, most of the urban single vehicle crashes included in the dataset are pedestrian hit (73 % of single vehicle crashes). Hence, the decrease in single vehicle crash likelihood can be attributed to the nature of transition areas, which normally exhibit low pedestrian volumes. Another interesting aspect of the results arises from the identified differences in the effect on angle crashes for segments and intersections (namely, notable decrease and increase in angle crash likelihood, respectively). In this case, the underlying crash mechanisms are most likely different: on transition segments, there is a considerable decrease in the number of driveways/minor intersections related to angle crashes, while the causes of angle crashes at intersections are still relevant and their likelihood was actually found to increase.

The "city centre" area type is influential for segments only and it is mainly related to an evident decrease in the angle crash likelihood. In this specific dataset, segments in the city centre are considerably short (i.e., on average, between 50 and 100 m long) and often configured as one-way roadways, in several cases single lane roadways with on-street parking on both sides. This may prevent reaching high speeds between two close intersections (see e.g. Silvano and Bang, 2015). Hence, drivers may experience possible angle conflicts without resulting in angle crashes.

Finally, concerning excluded variables, it is worth to note that the intersection type is not found to have a statistically significant effect on the likelihood of different crash types. This may seem contrary to expectations as the driving behaviour may significantly differ in signalized and unsignalized intersections (Liu, 2007; Li et al., 2019) and angle crashes are generally anticipated to decrease at intersections treated with traffic signals (see Jensen et al., 2010), even this effect may depend on several variables such as e.g., traffic volume ranges. However, on one hand, the number of entering lanes (included in the intersection model) can serve as a proxy variable for the intersection type (likely presence of traffic signals in case of several entering lanes) and complexity. On the other hand, during night, some of the traffic control systems may be not active, as such, their presence may not be influential on the safety performances. Moreover, there are instances where total crash frequencies of the two intersection types may be comparable for similar ranges of traffic volumes (see, for example, the models developed by Persaud et al., 2002), or the presence of traffic signals may not be influential for crash frequency predictions (Gomes et al., 2012).

The traffic volume for segments (contrary to the typical traffic which is significant), the segment length and the presence of bike paths are other not statistically significant determinants of crash type likelihood. The scarce influence of segment length may be due to the low variability of lengths in the dataset (see Table 1) or it may partially be captured by the area type variable. Finally, all bike paths in the sample are physically separated from the main roadway, thus explaining their scarce influence.

4.2 Associating crash-specific variables to urban crash types

Several crash-specific variables, either extracted from the crash dataset or inferred using the available data, were modelled to predict different urban crash type likelihoods. Among these variables, the presented models include: typical traffic and period of the day. A summary of their association to different crash type likelihoods is provided in Table 7, as based on the computed pseudo-elasticities in Tables 3 and 5.

Table 7. Summary of the association of crash-specific variables to different urban crash types T (Single Vehicle = SV, Angle = AN, Rear-End = RE, Sideswipe = SS) likelihood (S = Segments, I = Intersections)

Crash-specific variables	Change in Probability of each crash type*							
	Single vehicle (SV)		Angle (AN)		Rear-end (RE)		Sideswipe (SS)	
	S	I	S	I	S	I	S	I
Typical traffic – *No delays*^								
Typical traffic – *Some delays expected*^	+	+	-	-	+	+	+	+
Typical traffic –*Delayed*^		++		--		++		++
Period of the day – *Night*^	++	-	++	+	-	--	--	-

**The sign "+" reflects a positive effect (i.e., the specific crash type likelihood is increasing), while the sign "–" reflects a negative effect (i.e., the specific crash type likelihood is decreasing).*

^Numbers of + and - reflect the magnitude of the pseudo-elasticities (+/- for up to ± 50% change, ++/-- for a change included between ± 50% and ± 100%, +++/--- for more than ± 100% change).

The typical traffic at the crash day/hour was included in both intersection and segment models, with the attributes: some delays expected and delayed (only for intersections). In cases in which both delayed and with some delays expected typical traffic can be associated with different crash types (i.e., at intersections), their effect is consistent. In fact, for each crash type, changing from some delays expected to delayed traffic, the same effect is preserved (i.e., positive or negative) and amplified in case of delayed traffic (i.e., an effect of greater magnitude). In particular, a delayed traffic results in a notable decrease of the likelihood for angle crashes. This finding could be explained by the expected decrease of speed in delayed traffic conditions, which may prevent collisions between traffic streams having conflicting angles at intersections (see e.g., Wang et al., 2009).

The variable representing traffic with "some delays expected" would capture intermediate conditions in which there is neither free-flow traffic nor congestion. In such conditions, drivers are still likely to have some freedom in choosing speeds and trajectories according to their desires, but their choices could be constrained by the presence of other drivers. For intersections, as already stated, traffic with some delays expected was found to affect different crash type likelihoods similarly to the delayed traffic variable, even to a minor extent. Moreover, the different effects on crash types found for segments are similar to those discussed for the intersections..

Time-of-the-day when the crash occurred, and particularly, night time was also found to affect different crash type likelihoods at segments and intersections, but with substantial variations. A consistent reduction of rear-end and sideswipe crashes was identified for both segments and intersections during night. Rear-end crashes can be associated to high speeds (Islam, 2016), short headways and drivers' distraction (Gao and Davis, 2017). Under conditions of reduced visibility (even in the presence of lighting), it is likely that the driver would compensate for reduced visibility with a more cautious (Bella et al., 2014) and attentive behaviour. The highly attentive behaviour could result in promptly reacting to abrupt braking of preceding vehicles. Moreover, the intentions of drivers of the preceding vehicles can be more clear because of the increased visibility of car lights, compared to the daylight condition. The reduced likelihood of rear-end crashes at night is more evident at intersections (coherently with results from Yan et al., 2005), possibly because of even greater drivers' attention in cases of critical decision points such as intersections and the reduced number of vehicles with respect to daytime (Yan et al., 2005). Changing lanes may be particularly associated to segment-related sideswipe crashes (see Bham et al., 2012), as well as overtaking.. During nights, drivers may be more cautious when undertaking these types of manoeuvres on segments and, in fact, the reduced likelihood of sideswipe crashes at night is more evident at segments. An interesting difference stems from the indirect estimated effect of night-time on single-vehicle crashes: the latter are likely to decrease at intersections, but to increase at segments (in consistency with Bham et al., 2012). However, an increase in the angle night crashes likelihood was noted, which can be linked to lack of visibility for conflicting vehicles.

Seasonal and weekly variations are potentially related to different driving behaviour but also to different drivers' population (Intini et al., 2018), but they were not found significant for crash types. The influence on safety of seasonal and weekly variation may be more evident in rural than in urban areas, for instance because of the presence of summer/weekend recreational drivers (Intini et al., 2019c). Moreover, the effect of wet pavements may be more influential in rural rather than in urban environments (e.g. on run-off-road crashes, see McLaughlin et al., 2009). However, note that in the study by Bham et al. (2012), in which urban roadways were considered, weekends and wet pavements were associated to an increase in the single vehicle likelihood compared to other crash types.

4.3 Site-specific variability of estimated parameters

The random parameter model structure used in this study allows the identification of the variable effect of some predictors across the sites, based on the model estimates. As far as these predictors are concerned, the grouped random parameter structure enables the computation of a separate parameter estimate (β) corresponding to each individual segment/intersection. The variables that were found to have statistically significant grouped random parameters, and for which, segment- or intersection-specific parameters were estimated are (see also Tables 2 and 4):

- Period of the day (night: 6 p.m.-6 a.m.), in the sideswipe crash likelihood function for segments;
- Period of the day (night: 6 p.m.-6 a.m.), in the rear-end crash likelihood function for intersections;
- Traffic volume per entering lane, in the angle crash likelihood function for intersections.

Boxplots of the distribution of the three sets of parameters individually estimated for each site are reported in the next Figure for the sake of a thorough discussion about their variability. The distributions of the individually estimated parameters were taken into account, rather than the computed distributions based on the estimated means and standard deviations, as the former lead to higher forecasting accuracy according to previous research (Anastasopoulos, 2016: Fountas and Anastasopoulos, 2017; Fountas et al., 2018b).

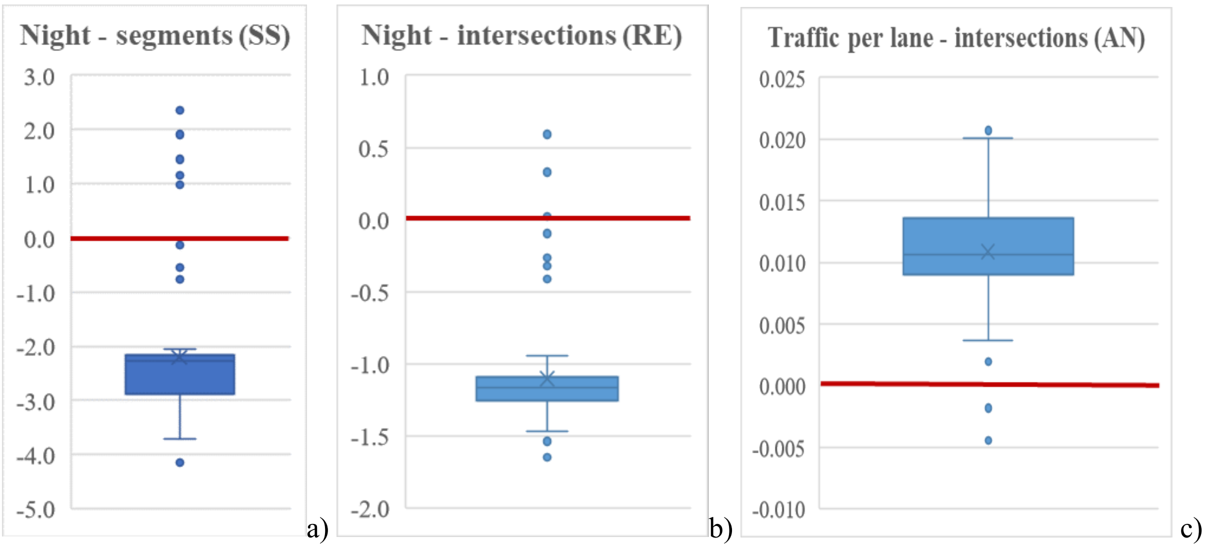


Figure 2. Boxplots of the distributions of the three grouped random parameters (with boxes delimiting the interquartile range IQR = Q_3,75th – Q_1,25th, whiskers at 1.5 times the IQR in both directions and solid lines indicating the 0 value). Parameter distributions from left to right: a) period of the day (night), sideswipe crashes - segment model; b) period of the day (night), rear-end crashes - intersection model; c) traffic per lane, angle crashes - intersection model.

The distribution of coefficients varies depending on the associated explanatory variable; specifically, the boxplots show a considerably broad range for the night variable, especially for segments, and a small range of variation for the traffic variable. All the distributions of estimated parameters in Figure 2 have some "outliers" (conventionally identified as above or below 1.5 times the interquartile range of the distribution). However, it is crucial to note that the effect of a given variable is generally positive/negative for all the segments/intersections, except for some of these outliers, where the effect is reversed. Those cases are discussed in the following.[1]

For what concerns the night effect in the segment model, it is directly related to a decrease in the sideswipe crash likelihood for 110 segments (92 % of the population). However, for 9 segments (8 % of the population), positive parameters were estimated. An investigation of the characteristics of these segments has revealed that most of them are undivided roads with parked vehicles on both sides (in some cases coupled with narrow lanes and one-way traffic). The mechanism of sideswipe crashes can be eased by the presence of side parking on narrow roads or in cases of roads with more-than-one lanes, by possible lane change and overtaking manoeuvres, especially at night. These situations are actually likely to occur in most of the highlighted sites showing positive parameter estimates.

In contrast, the night effect in the intersection model is directly related to a decrease in the rear-end crashes for 125 intersections (97 % of the population). However, for 4 intersections (3 % of the population), the parameter estimates were found to be positive. Two out of these four intersections consist of a major arterial road, which intersect a minor road. The presence of a high-volume road may foster rear-end crashes, because high speeds are operated and abrupt braking may occur at intersections, especially in low visibility conditions. On the other hand, the other two intersections are four-legged signalized intersections with unbalanced traffic between the major and the minor road (especially in one case). In these cases, it is possible that with the lower night-time traffic, drivers on the main road may operate higher speeds as well, fostering the same mechanism of abrupt braking at the signalized intersection with the minor road (whether it is normally working or with flashing lights at night) and the related rear-end mechanism.

For what concerns the effect of traffic volume in the intersection model, an increase in the mean traffic volume per entering lane is directly related to an increase in the likelihood of angle crashes on 126 intersections (98 % of the population), likely due to the increased angular conflicts. However, there are three intersections (2 % of the sample) for which the traffic volume parameter estimate is negative. In one intersection, there is one major two-way two-lane road and a one-way minor road, on which the traffic from the major road can only enter into. Hence, in this case, angle crashes could be only caused by the left-turn manoeuvre from the major to the minor road. As the traffic volume increases, drivers may be more cautious while negotiating the left-turn manoeuvre; the risk compensating behaviour of drivers in such cases may explain the reduction in the angle crash likelihood. In another case, the intersection is between an entering one-way road and a major two-lane road, having an angle greater than 90°. In this case, the vehicle flow from the minor road (give-way regulated) enters almost parallel to the direction of vehicles on the main road. In fact, half crashes on this site are sideswipe crashes. Hence, in this case, the effect of traffic on angle crashes is not influential. The third case is a four-legged signalized intersection with highly unbalanced traffic between the major and the minor road. In this case, angular conflicts are largely independent on the average traffic per lane (mainly governed by the main road traffic). The most frequent crash type on this intersection is the rear-end crash indeed.

[1] Due to the five-year period of the crash data, there exists the possibility that some of the unobserved effects captured by the random parameters may stem from the temporal instability of factors affecting the crash types. The effect of temporal instability on statistical modelling of crash data has been extensively discussed by Mannering, 2018; Almawasi and Mannering, 2019; Behnood and Mannering, 2019.

5. Practical application of results

The estimated models can be used in practice to highlight high-risk sites with respect to a given crash type. In fact, based on the models and the dataset, individual probabilities of occurrence of crash type outcomes can be assessed. In the estimated models (for segments and intersections), some site-specific (segment- or intersection-related) and crash-specific variables were included (see Fig. 3).

In this case, the high-risk sites identification should be aimed at highlighting sites having a very high probability of a specific crash type to occur. This procedure is carried out for particular combinations of crash-specific variables (which can be seen as crash contributing factors), leading to different possible scenarios. The criteria used to generate scenarios for both segments and intersections are shown in Fig. 3.

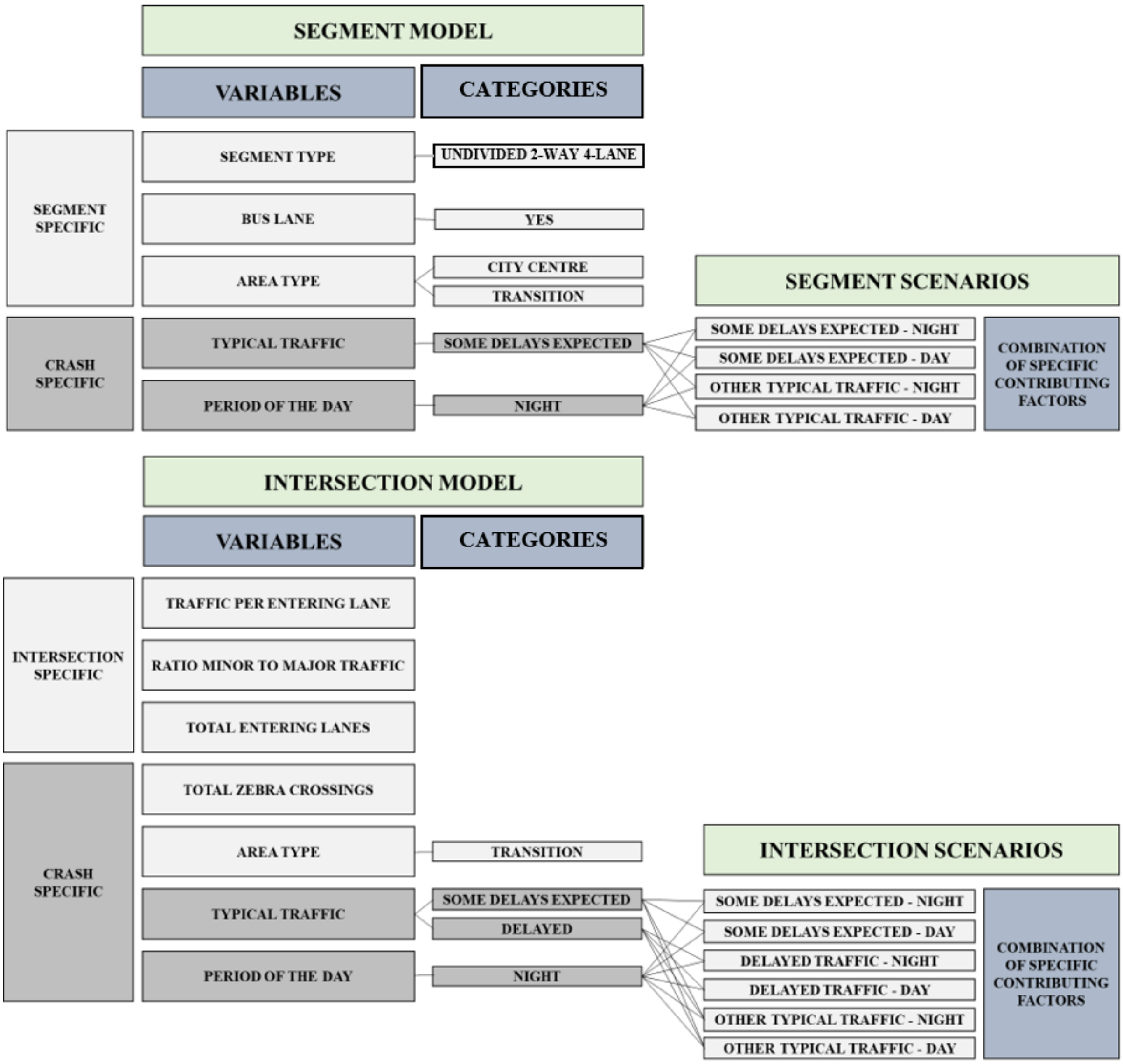
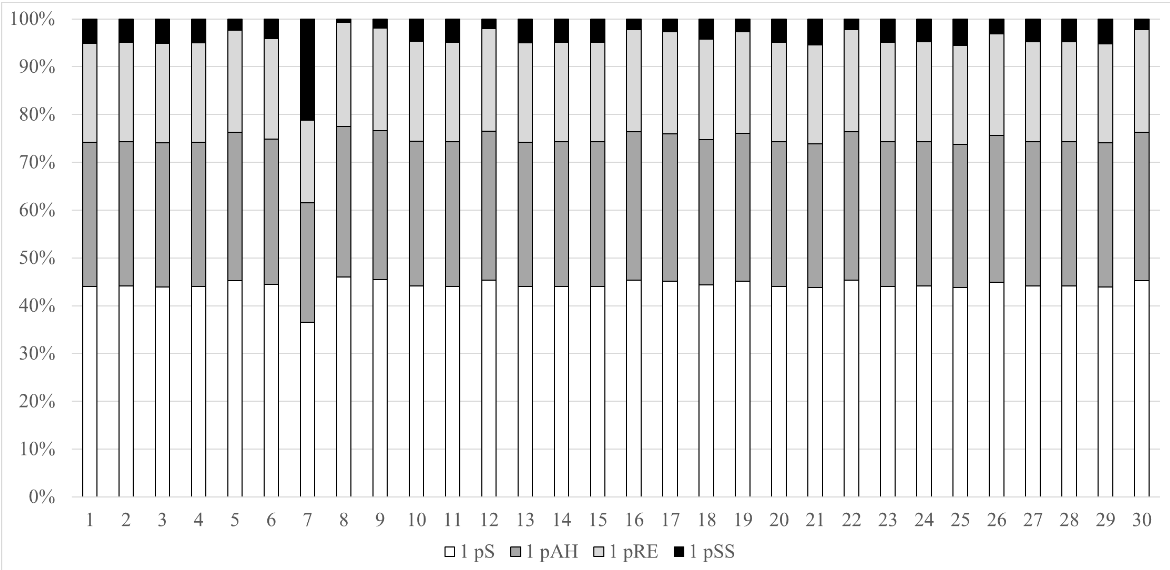
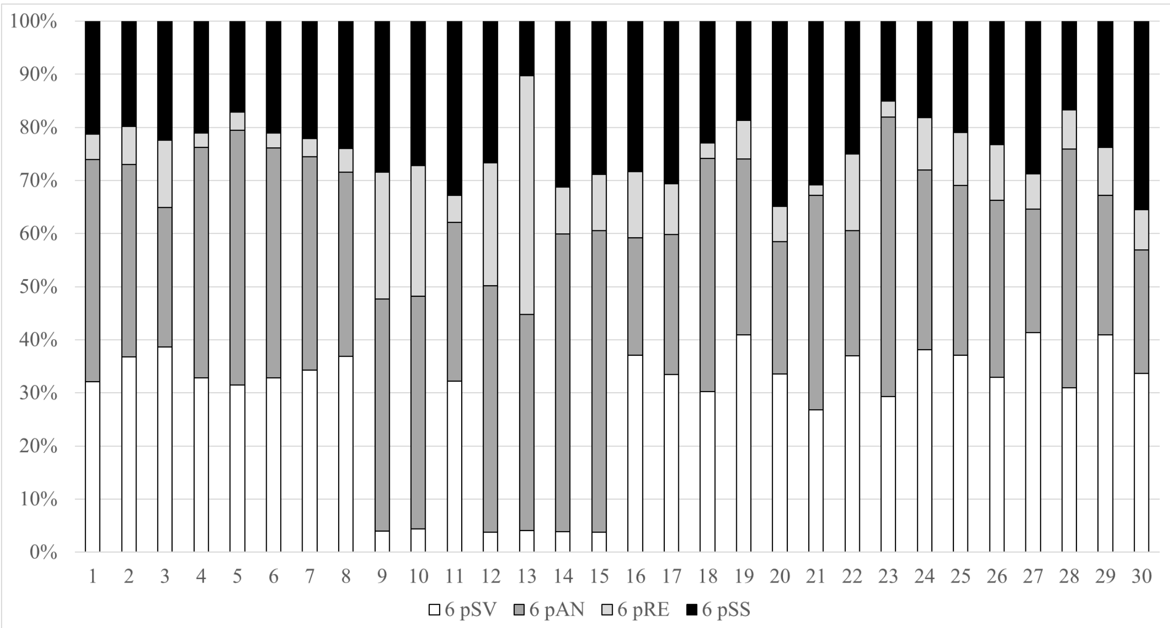


Figure 3. Generation of the different scenarios for the high-risk sites identification, based on combinations of specific contributing factors

In detail, the probabilities associated to different crash types were computed in four different scenarios for segments, and six different scenarios for intersections, as indicated in Figure 3. The four segment scenarios are: traffic with some delays expected/day, traffic with some delays expected/night, other traffic conditions different than some delays expected/night, other traffic conditions different than some delays expected/day. The six intersection scenarios are: delayed traffic/night, delayed traffic/day, traffic with some delays expected/night, traffic with some delays expected/day, no delays expected (or unavailable data for typical traffic)/night, no delays expected (or unavailable data for typical traffic)/day. The practical meaning of the identified scenarios lies in the possibility of computing different crash type likelihoods for different conditions. For instance, different likelihoods are associated with the delayed traffic in both the day and night periods, which may reflect, namely, the morning peak hour, and the afternoon peak hour. Some examples of the crash type probability distributions are provided in Figure 4 for both segments and intersections.



a)



b)

Figure 4. a) Examples of crash type T (Single Vehicle = SV, Angle = AN, Rear-End = RE, Sideswipe = SS) probability *p* distribution for the samples of segments (in the example scenario 1: night-traffic with some delays expected). b) Examples of crash type T probability distribution *p* for the samples of intersections (in the example scenario 6: night-delayed traffic). Sub-sets of 30 sites only are used for illustrative purposes in both plots.

Based on this approach, high-risk sites having high likelihood of a given crash type to occur, can be identified in the different scenarios for both segments and intersections, by setting given thresholds depending on the scope of high-risk sites analysis. For example, starting from the population of all the computed probabilities of different crash types for all sites (segments or intersections), it is possible to define some threshold percentiles (e.g., 85th, 90th or 95th percentile). The definition of thresholds may depend on the scope of the analysis (exploratory purposes, network screening, inspection planning, etc.). Once thresholds are defined, the sites showing percentages of crashes of a given type exceeding the thresholds, can be identified as "high-risk sites" for that crash type. This detailed analysis may result in selecting countermeasures specifically related to given crash types.

6. Conclusions

In this study, a dataset of urban segments and intersections was used to identify the factors influencing the likelihood of different crash types (single-vehicle, angle, rear-end and sideswipe). A multinomial logit approach, with different crash types serving as outcomes and several traffic, geometric and context-related variables serving as possible explanatory variables, was implemented. In detail, the mixed model structure was used to account for the variability of estimates across the crash observations. Parameter estimates were grouped per road site (segment/intersection), in order to account for unobserved effects and assess the influence of predictors on crash types at the individual site level, which is a research novelty for crash type modelling to the authors' knowledge, especially for urban crashes. The main aim of this study was to explore: a) the influence of geometric and traffic-related predictors on different urban crash types (both at segments and intersections); b) the association of crash-specific variables to urban crash types, c) the possible variability of results across the crash observations for individual segments and intersections.

The results show that the segment type and the presence of bus lanes are predictors of different types of crash occurring on road segments. Traffic volume per entering lane, total number of entering lanes, total number of zebra crossings and the ratio between major and minor traffic volumes at intersections influence different crash types at intersections. The context variable: area type is a predictor of different crash types for both urban segments and intersections.

The crash-specific variables, which were significantly associated with different crash types (for both segments and intersections), are the typical traffic at the moment of the crash and the period of the day. However, no significant seasonal and weekly variations were noted, as well as no influence of different pavement conditions. It is important to note that a measure of the traffic conditions at the moment of the crash (even if inferred from online sources) was statistically associated with different crash types. Hence, the use of similar variables is encouraged for future research.

For the predictors associated to statistically significant grouped random parameters (period of the day for both segments and intersections, traffic volume per entering lane), substantial variability of their effect was identified across the crash observations. Occasionally, the direction of the effects of some variables is the opposite of what holds to all the other elements in the population. In these cases, the further analyses conducted on these particular sites have revealed the influence of some local factors on the estimation of the parameters with different sign. The disclosure of possible local relationships constitutes a direct implication of the grouped random parameter approach and corroborates the choice of such approach. In fact, differently than in the conventional mixed logit, the grouped random parameter approach can capture not only unobserved effects varying across the crash population, but also systematic variations arising from the unobserved interaction between the geometric or traffic characteristics of these sites and the drivers' behavioural response against them (Fountas et al., 2018b). In addition, the estimation of individual parameters can help better identify the potential sources of these unobserved interactions at a segment or intersection level.

Hence, this study contributes to the existing body of research since it is the first to show, to the authors' knowledge, how the grouped random parameter multinomial logit structure can be implemented to account for unobserved and grouped heterogeneity in crash type prediction. The introduction of the grouped random parameters to the multinomial logit formulation constitutes a significant comparative advantage of the presented models relative to state-of-practice approaches. In fact, the presented approach allows for capturing the impact of unobserved factors that may vary across the segments/intersections (i.e., unobserved heterogeneity) as well as grouped effects arising from the presence of multiple crash observations per segment or intersection. Over the last few years, the impact of segment- or intersection-specific grouped heterogeneity has been recognized in various safety dimensions, such as the accident occurrence (Fountas et al. 2018b) or the injury severity (Fountas et al., 2018a); however, the implications of grouped heterogeneity on crash type probability have not been thoroughly explored to date. It should be noted that the formulations of SPFs or other state-of-practice modeling approaches do not typically take into account unobserved or grouped heterogeneity, hence resulting in less accurate parameter estimates and statistical inferences (Washington et al., 2020).

Moreover, the results from the empirical analysis can be practically used to highlight high-risk segments or intersections with specific regard to given crash type outcomes, differentiated by particular scenarios (obtained as combinations of contributing factors, as for example, specific time of the day or traffic conditions). This can be considered as a step forward for the selection of appropriate and individual countermeasures at sites, based on their predicted crash type outcomes and considering other influential conditions.

The present study is not without limitations. Firstly, as most of research in road safety, the transferability of the estimated models to other contexts requires further investigation. Secondly, the sample size used for this study was deemed large enough for the exploratory purposes of this research, but it should be enlarged for prediction purposes. Moreover, several other variables (i.e., related to human factors or the role of vulnerable road users) may affect the crash types. However, the employed grouped random parameter approach can account for this limitation to a reasonable extent (Mannering et al., 2016). Note that even incorporating year specific effects in the discrete outcome models may add further value to this modelling approach, which could be considered for further research. Nevertheless, since the grouped random parameters follow pre-determined distributions, the practical application of these models is not as straightforward as in cases of more parsimonious models (such as the SPFs), where the parameter estimates have fixed values regardless of the characteristics of the segment/intersection. However, this limitation of the grouped random parameters models stems from their generalized formulation, which has been set to account for various layers of heterogeneity. As a concluding note, given the exploratory nature of this study, further research should deepen these findings, by possibly using larger datasets and different contexts, in order to compare results.

ACKNOWLEDGEMENTS

This study was carried out within the activities of the National Research Project Pa.S.S.S. (Parco Scientifico della Sicurezza Stradale, in English: Scientific Park for Road Safety), granted by the Italian Ministry of Transport and Infrastructures and funded by the Municipality of Bari, main scientific partner: Polytechnic University of Bari. All those involved in project tasks and preparatory related activities are then acknowledged. The Department of Civil, Environmental, Land, Building Engineering and Chemistry, Polytechnic University of Bari, Italy, is acknowledged for having partially funded the stay of the first author at the Edinburgh Napier University, Scotland (UK), essential for the realization of this article.

REFERENCES

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention*, 38(2), 215-224.

Aarts, L., Commandeur, J., Welsh, R., Niesen, S., Lerner, M., Thomas, P., Bos, N., Davidse, R. (2016). *Study on Serious Road Traffic Injuries in the EU*. European Commission. Brussels.

AASHTO (American Association of State Highway and Transportation Officials) (2010). *Highway Safety Manual*; Washington, DC.

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, 34(5), 597-603.

Ackeret, K., Nambisan, S., & Menon, R. (1999). Evaluation of Side Swipe Crashes at Triple and Dual Left Turn Lanes in Las Vegas, Nevada. In *Transportation Frontiers for the Next Millennium: 69th Annual Meeting of the Institute of Transportation Engineers* Institute of Transportation Engineers (ITE) (No. Publication No. CD-006).

Alnawmasi, N. and Mannering, F. (2019). A statistical assessment of temporal instability in the factors determining motorcyclist injury severities. *Analytic Methods in Accident Research*, 22, 100090.

Anastasopoulos, P. (2016). Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic Methods in Accident Research*, 11, 17-32.

Bauer, R., Machata, K., Brandstaetter, C., Yannis, G., Laiou, A., & Folla, K. (2016). Road traffic accidents in European urban areas. In *1 st European Road Infrastructure Congress*, Leeds.

Bella, F., Calvi, A., & D'Amico, F. (2014). Analysis of driver speeds under night driving conditions using a driving simulator. *Journal of Safety Research*, 49, 45.e1-52.

Bham, G., Javvadi, B., & Manepalli, U. (2011). Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban US highways in Arkansas. *Journal of Transportation Engineering*, 138(6), 786-797.

Behnood, A. and Mannering, F. (2019). Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. *Analytic Methods in Accident Research*, 23, 100102.

Bhowmik, T., Yasmin, S., & Eluru, N. (2019). Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types. *Analytic Methods in Accident Research*, 24, 100107.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., & Wang, X. (2018). Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic Methods in Accident Research*, 19, 1-15.

Celik, A., & Oktay, E. (2014). A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. *Accident Analysis and Prevention*, 72, 66-77.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., & Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. *Accident Analysis and Prevention*, 80, 76-88.

Chen, Y., Wang, K., King, M., He, J., Ding, J., Shi, Q., Wang, C. & Li, P. (2016). Differences in factors affecting various crash types with high numbers of fatalities and injuries in China. *PLoS one*, 11(7), e0158559.

Christoforou, Z., Cohen, S., & Karlaftis, M. (2011). Identifying crash type propensity using real-time traffic data on freeways. *Journal of Safety Research*, 42(1), 43-50.

Colonna, P., & Berloco, N. (2011). *Transition zones between urban and inter-urban areas: Identification and definition criteria*. In 24th World Road Congress World Road Association (PIARC). Mexico City.

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

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

Colonna, P., Berloco, N., Cavalluzzi, G., Fedele, V., Intini, P., & Ranieri, V. (2019b). *Pa.S.S.S. Research Project, Second Year Report*. Politecnico di Bari.

Eker, U., Ahmed, S., Fountas, G. & Anastasopoulos, P. (2019). An exploratory investigation of public perceptions towards safety and security from the future use of flying cars in the United States. *Analytic Methods in Accident Research*, 23, 100-103.

Elvik, R. (2013). A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. *Accident Analysis and Prevention*, 50, 854-860.

Eurostat Data. Available online: https://ec.europa.eu/eurostat.

Fountas, G., & Anastasopoulos, P. (2017). A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. *Analytic Methods in Accident Research*, 15, 1-16.

Fountas, G., Anastasopoulos, P. & Mannering, F. (2018a). Analysis of vehicle accident-injury severities: a comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. *Analytic Methods in Accident Research*, 18, 15-32.

Fountas, G., Sarwar, M., Anastasopoulos, P., Blatt, A., & Majka, K. (2018b). Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. *Accident Analysis and Prevention*, 113, 330-340.

Fountas, G., Pantangi, S., Hulme, K., Anastasopoulos, P. (2019). The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach. *Analytic Methods in Accident Research*, 22, 100091.

Gao, J., & Davis, G. (2017). Using naturalistic driving study data to investigate the impact of driver distraction on driver's brake reaction time in freeway rear-end events in car-following situation. *Journal of Safety Research*, 63, 195-204.

Geedipally, S., Patil, S., & Lord, D. (2010). Examination of methods to estimate crash counts by collision type. *Transportation Research Record: Journal of the Transportation Research Board*, 2165, 12-20.

Gomes, S.; Geedipally, S.; Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety Science*, 50, 1732–1739.

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

Halton, J. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1), 84-90.

Harwood, D.; Bauer, K.; Richard, K.; Gilmore, D.; Graham, J.; Potts, I.; Torbic, D.; Hauer, E. (2007). *Methodology to Predict the Safety Performance of Urban and Suburban Arterials*; No. NCHRP Project 17-26; National Academy of Sciences: Washington, DC.

Heydari, S., Fu, L., Thakali, L. & Joseph, L. (2018). Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: Applications to grade crossing safety analysis. *Analytic Methods in Accident Research*, 19, 33-48.

Hole, A. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, 7(3), 388-401.

Intini, P., Berloco, N., Colonna, P., Ranieri, V., & Ryeng, E. (2018). Exploring the relationships between drivers' familiarity and two-lane rural road accidents. A multi-level study. *Accident Analysis and Prevention*, 111, 280-296.

Intini, P., Berloco, N., Binetti, R., Fonzone, A., Ranieri, V., & Colonna, P. (2019a). Transferred versus local Safety Performance Functions: A geographical analysis considering two European case studies. *Safety Science*, 120, 906-921.

Intini, P., Berloco, N., Cavalluzzi, G., Lord, D., Ranieri, V., Colonna, P. (2019b). *The variability of urban Safety Performance Functions with different road elements: an Italian case study*. Paper accepted for presentation at the 99th Annual Meeting of the Transportation Research Board 2020, Washington, DC.

Intini, P., Berloco, N., Colonna, P., Ottersland Granås, S., & Olaussen Ryeng, E. (2019c). Influence of Road Geometric Design Consistency on Familiar and Unfamiliar Drivers' Performances: Crash-Based Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2673, 489-500.

Islam, M. (2016). Investigating the speed and rear-end collision relationship at urban signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2601, 10-16.

Jensen, S. (2010). Safety effects of intersection signalization: A before-after study. In *TRB 89th Annual Meeting Compendium of Papers CD-ROM*. Washington, DC.

Jonsson, T., Ivan, J., & Zhang, C. (2007). Crash prediction models for intersections on rural multilane highways: Differences by collision type. *Transportation Research Record: Journal of the Transportation Research Board*, 2019, 91-98.

Jonsson, T., Lyon, C., Ivan, J., Washington, S., van Schalkwyk, I., & Lord, D. (2009). Differences in the performance of safety performance functions estimated for total crash count and for crash count by crash type. *Transportation Research Record: Journal of the Transportation Research Board,* 2102, 115-123.

Kim, D., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis and Prevention*, 39(1), 125-134.

Kim, D., Washington, S., & Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering*, 132(4), 282-292.

Kim, J., Ulfarsson, G., Kim, S., & Shankar, V. (2013). Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis and Prevention*, 50, 1073-1081.

Kockelman, K., & Kweon, Y. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis and Prevention*, 34(3), 313-321.

Lantieri, C., Lamperti, R., Simone, A., Costa, M., Vignali, V., Sangiorgi, C., & Dondi, G. (2015). Gateway design assessment in the transition from high to low speed areas. *Transportation Research Part F*, 34, 41-53.

Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., & Cai, Q. (2018). Analysis of crash proportion by vehicle type at traffic analysis zone level: A mixed fractional split multinomial logit modeling approach with spatial effects. *Accident Analysis and Prevention*, 111, 12-22.

Li, G., Wang, Y., Zhu, F., Sui, X., Wang, N., Qu, X., & Green, P. (2019). Drivers' visual scanning behavior at signalized and unsignalized intersections: A naturalistic driving study in China. *Journal of Safety Research*, 71, 219-229.

Lin, C., Chuang, C., Kerick, S., Mullen, T., Jung, T., Ko, L., Chen, S., King, J., & McDowell, K. (2016). Mind-wandering tends to occur under low perceptual demands during driving. *Scientific Reports*, 6, 21353.

Liu, B. (2007). Association of intersection approach speed with driver characteristics, vehicle type and traffic conditions comparing urban and suburban areas. *Accident Analysis and Prevention*, 39(2), 216-223.

Lord, D.; Persaud, B. (2000). Accident prediction models with and without trend: Application of the generalized estimating equations procedure. *Transportation Research Record: Journal of the Transportation Research Board*, 1717, 102–108.

Malyshkina, N., & Mannering, F. (2009). Markov switching multinomial logit model: an application to accident-injury severities. *Accident Analysis and Prevention*, 41(4), 829-838.

Mannering, F., Shankar, V., & Bhat, C. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, 11, 1-16.

Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, 17, 1-13.

McLaughlin, S., Hankey, J., Klauer, S., & Dingus, T. (2009). *Contributing factors to run-off-road crashes and near-crashes*. United States Department of Transportation. National Highway Traffic Safety Administration. Report No. DOT HS 811 079.

Milton, J., Shankar, V., & Mannering, F. (2008). Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis and Prevention*, 40(1), 260-266.

Moore, D., Schneider IV, W., Savolainen, P., & Farzaneh, M. (2011). Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident Analysis and Prevention*, 43(3), 621-630.

Mothafer, G., Yamamoto, T. & Shankar, V. (2016). Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. *Analytic Methods in Accident Research*, 9, 16-26.

NHTSA (National Highway Traffic Safety Administration), US Department of Transportation (2019). *Traffic Safety Facts*. *Rural/Urban Comparison of Traffic Fatalities*. DOT HS 812 741.

Pantangi, S., Fountas, G., Sarwar, M., Anastasopoulos, P., Blatt, A., Majka, K., Pierowicz, J. & Mohan, S. (2019). A preliminary investigation of the effectiveness of high visibility enforcement programs using naturalistic driving study data: a grouped random parameters approach. *Analytic Methods in Accident Research*, 21, 1-12.

Persaud, B.; Lord, D.; Palmisano, J. (2002). Calibration and transferability of accident prediction models for urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 1784, 57–64.

Retting, R., Williams, A., Preusser, D., & Weinstein, H. (1995). Classifying urban crashes for countermeasure development. *Accident Analysis and Prevention*, 27(3), 283-294.

Russo, B., Savolainen, P., Schneider IV, W., & Anastasopoulos, P. (2014). Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. *Analytic Methods in Accident Research*, 2, 21-29.

Sacchi, E., Persaud, B., & Bassani, M. (2012). Assessing international transferability of highway safety manual crash prediction algorithm and its components. *Transportation Research Record: Journal of the Transportation Research Board*, 2279, 90-98.

Sarwar, M., Anastasopoulos, P., Golshani, N., & Hulme, K. (2017). Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: a driving simulation study. *Analytic Methods in Accident Research*, 13, 52-64.

Savolainen, P., 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.

Savolainen, P., & Mannering, F. (2007). Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. *Accident Analysis and Prevention*, 39(5), 955-963.

Sayed, T.; Rodriguez, F. (1999). Accident prediction models for urban unsignalized intersections in British Columbia. *Transportation Research Record: Journal of the Transportation Research Board*, 1665, 93–99.

Shankar, V., & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of safety research*, 27(3), 183-194.

Shi, Q., & Abdel-Aty, M. (2015). Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C*, 58, 380-394.

Silvano, A., & Bang, K. (2015). Impact of speed limits and road characteristics on free-flow speed in urban areas. *Journal of Transportation Engineering*, 142(2), 04015039.

StataCorp. (2015). *Stata: Release 14. Statistical Software*. College Station, TX: StataCorp LLC.

Sun, D., & Elefteriadou, L. (2012). Lane-Changing Behavior on Urban Streets: An "In-Vehicle" Field Experiment-Based Study. *Computer-Aided Civil and Infrastructure Engineering*, 27(7), 525-542.

Tay, R., Choi, J., Kattan, L., & Khan, A. (2011). A multinomial logit model of pedestrian–vehicle crash severity. *International Journal of Sustainable Transportation*, 5(4), 233-249.

Wali, B., Khattak, A., & Khattak, A. (2018). A heterogeneity based case-control analysis of motorcyclist's injury crashes: Evidence from motorcycle crash causation study. *Accident Analysis and Prevention*, 119, 202-214.

Wang, X., Abdel-Aty, M., Almonte, A., & Darwiche, A. (2009). Incorporating traffic operation measures in safety analysis at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2103, 98-107.

Washington, S., Karlaftis, M., Mannering, F. & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. CRC Press, Taylor and Francis Group, New York, NY.

Werneke, J., & Vollrath, M. (2012). What does the driver look at? The influence of intersection characteristics on attention allocation and driving behavior. *Accident Analysis and Prevention*, 45, 610-619.

Wu, Q., Chen, F., Zhang, G., Liu, X., Wang, H., & Bogus, S. (2014). Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. *Accident Analysis and Prevention*, 72, 105-115.

Yan, X., Radwan, E., & Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis and Prevention*, 37(6), 983-995.

Yasmin, S. & Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis and Prevention*, 59, 506-521.

Yasmin, S., Eluru, N., Pinjari, A. & Tay, R. (2014). Examining driver injury severity in two vehicle crashes–A copula based approach. *Accident Analysis and Prevention*, 66, 120-135.