Development of Macro-Level Safety Performance Functions in the City of Naples

Alfonso Montella [1], Vittorio Marzano [1], Filomena Mauriello [1], Roberta Vitillo [2], Roberto Fasanelli [3], Mariano Pernetti [4] and Francesco Galante [1,*]

[1] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, 80125 Naples, Italy; alfonso.montella@unina.it (A.M.); vmarzano@unina.it (V.M.); filomena.mauriello@unina.it (F.M.)

[2] Department of Business and Quantitative Studies, University of Naples Parthenope, 80132 Naples, Italy; roberta.vitillo@uniparthenope.it

[3] Department of Social Sciences, University of Naples Federico II, 80138 Naples, Italy; roberto.fasanelli@unina.it

[4] Department of Engineering, University of Campania "Luigi Vanvitelli", 81031 Aversa (Caserta), Italy; mariano.pernetti@unicampania.it

* Correspondence: francesco.galante@unina.it; Tel.: +39-081-768-3940

Abstract: This paper presents macro-level safety performance functions and aims to provide empirical tools for planners and engineers to conduct proactive analyses, promote more sustainable development patterns, and reduce road crashes. In the past decade, several studies have been conducted for crash modeling at a macro-level, yet in Italy, macro-level safety performance functions have neither been calibrated nor used, until now. Therefore, for Italy to be able to fully benefit from applying these models, it is necessary to calibrate the models to local conditions. Generalized linear modelling techniques were used to fit the models, and a negative binomial distribution error structure was assumed. The study used a sample of 15,254 crashes which occurred in the period of 2009–2011 in Naples, Italy. Four traffic analysis zones (TAZ) levels were used, as one of the aims of this paper is to check the extent to which these zoning levels help in addressing the issue. The models were developed by the stepwise forward procedure using explanatory Socio-Demographic (S-D), Transportation Demand Management (TDM), and Exposure variables. The most significant variables were: children and young people placed in re-education projects, population, population aged 65 and above, population aged 25 to 44, male population, total vehicle kilometers traveled, average congestion level, average speed, number of trips originating in the TAZ, number of trips ending in the TAZ, number of total trips and, number of bus stops served per hour. An important result of the study is that children and young people placed in re-education projects negatively affects the frequency of crashes, i.e., it has a positive safety effect. This demonstrates the effectiveness of education projects, especially on children from disadvantaged neighbourhoods.

Keywords: safety prediction models; traffic analysis zone; negative binomial; traffic crashes

1. Introduction

Road safety has been increasingly regarded as one of the most important transportation concerns in urban areas. Over the last few decades, the development of safety performance functions has enabled traffic engineers and road safety researchers to identify important factors related to the occurrence of crashes on specific highway elements or on transportation networks [1].

A safety performance function (SPF) is an equation used to predict the average number of crashes per year at a location as a function of exposure and, in some cases, it includes site characteristics [2].

This type of model belongs to a family of generalized linear models (GLM) with a non-normal error structure distribution [3].

The SPFs can be classified into two different spatial aggregation scales: micro-level and macro-level. In the first scale, the study units are based on small homogeneous road entities, such as roadway segments, ramps, and intersections [4,5]. The micro-level factors refer to variables aggregated at the segment/intersection level including traffic data and, geometric data (e.g., number of lanes, road functional classification). In macro-level analysis, the study units are based on some geographic areas (zonal-level) to investigate the influence of socio-economic, demographic, land use, and infrastructure-related factors on crash occurrence [4]. Several studies have been conducted for crash modelling at a macro-level, exploring various zonal systems: block groups, census tracts, ZIP code areas and, traffic analysis zones (TAZs) [6–13]. Most of these zonal systems were developed for different specific uses. The prevalent spatial unit considered at the macro-level analysis is TAZ. A TAZ may consist of one or more census blocks, block groups, or census tracts; but usually it is a spatial aggregation of census blocks. TAZ boundaries generally coincide with identifiable physical barriers such as major streets and water bodies, and they are delineated in such a way that within each TAZ the land use activities are relatively homogeneous [7].

The objective of these models is to establish relationships between the number of crashes per traffic analysis zone and neighbourhood traits (explanatory variables), such as traffic, road network characteristics, socioeconomic and demographic features, land use, dwelling unit, and employment type. Macro-level safety performance functions that are consistent with aggregate travel demand models have been developed to provide empirical tools for planners and engineers to conduct proactive analyses, promote more sustainable development patterns, and reduce the road crash burden on communities worldwide [1–37]. These models have great potential to promote increasingly sustainable development patterns by combining several redeeming features from pre-existing models. Specifically, improvements in land use and infrastructure efficiency, a reduction in environmental impact, an increased walkability and an improved neighbourhood social environment [38].

Some of the dependent variables modeled in the previous studies include: total crashes; severe injury crashes; peak morning crashes; property damage only (PDO) crashes; total number of fatalities; total number of injuries; pedestrian crashes and; number of crashes involving elderly drivers [7,11,20,39,40].

Explanatory variables used in previous studies can be grouped in four classes [15]: (1) traffic characteristics (Exposure), (2) social demographic factors (SD), (3) roadway factors (Network), and (4) land use and travel habits (Transportation Demand Management).

Although it is not the most significant predictor of crashes, exposure is a key determinant of traffic safety. The relationship between crash occurrence and exposure is fairly straightforward. The higher the exposure, the greater the possibility for a crash to occur [18]. The most common exposure variables, annual average daily traffic (AADT) or vehicle kilometers traveled (VKT), were used along with average zonal operating speed (SPD), and average zonal volume to capacity ratio (VC). Hadayeghi et al. [7] found that VKT had significant effects on crash occurrence in a nonlinear relationship. Lovegrove et al. [15,16] confirmed earlier research regarding the dominant influence of VKT on crash predictions of all types. It also highlighted the significant influence that congestion (VC) and average zonal operating speed (SPD) play in safety evaluation.

Several studies observed that a low socioeconomic status and deprivation increase the fatality risk or the risk of being injured in traffic [17–19]. An area's socioeconomic deprivation level is usually measured by proxy factors such as total population (TPOP), population density (POPD), household density (NHD), and percentage employed (EMPP) within the TAZ [9]. Ladron de Guevara et al. [20] observed that population density and the number of employees (employment density) played a significant role in predicting crashes [18]. Lee et al. [24] also observed that a lower proportion of households without an available vehicle within a ZIP code was negatively associated with the risk of pedestrians being involved in a crash. Several authors have suggested factors including age,

and sex to explain crash risk [25]. Wier et al. [26] have shown that the proportion of the population living in poverty, and the number of people aged 65 and older as percentage of the total population, were significantly good predictors of crashes. Similarly, Ukkusuri et al. [27] found that the proportion of the uneducated (without any schooling) population had a positive effect on pedestrian crashes, while Lascala et al. [28] concluded that the proportion of high school graduates was inversely correlated with pedestrian injury collisions.

Several road networks factors were considered in macroscopic studies, such as zonal lane kilometers (TLKM), percentage of each road class (ALKMP, LLKMP), intersection density (INTD), signal density (SIGD), intersection type (I3WP, IALP), and the average curvature of roadways (CRVD). Some studies have shown, that roadway density has a positive association with total crashes [19] and fatal crashes [12]. Hadyeghi et al. [21] and Gomes et al. [22] observed that intersection density, number of households, the number of major road kilometres and, the number of vehicle kilometers traveled, all had significant effects on crash occurrence. Cai et al. [6] found that the length of sidewalks and length of bike lanes, have a positive effect on crash frequency.

Transportation Demand Management (TDM) strategies have hardly ever been implemented to improve traffic safety. Their main objectives are usually the reduction of congestion and emission, as well as travel costs and energy by means of reducing travel demand and consequently vehicle distance traveled, although their impact on traffic safety should not be neglected [29–31]. However, different individual daily trips and land use are examined in numerous crash investigations. Wedagama et al. [8] found that residential population density, manufacturing, retail trade and services industries were positively related to the number of road traffic crashes. Kim and Yamashita [41] observed that areas with mixed residential and commercial land use have a higher frequency of crashes. Moreover, Pulugurtha et al. [23] also observed that land use characteristics such as urban residential and mixed-use development are strongly associated with the number of crashes in a TAZ.

SPFs are critical to local and state transportation agencies due to their ability to identify regions with potential safety concerns [42]. Therefore, for a jurisdiction or nation to fully benefit from applying these models, it is necessary to calibrate or recalibrate them to local conditions [43]. This is because crash occurrence frequency, and the associated under- and over-dispersion in crash data can vary significantly across an area. The need for calibrating SPFs to specific area is clearly recognized by the American Association of State Highway and Transportation Officials (AASHTO) due to variations in factors associated with safety, such as road geometry and conditions, environmental factors, geographic characteristics, crash characteristics, reporting thresholds, all of which can be unique to a specific area [2,42].

Since macro-level has not been calibrated nor used until now in Italy, this paper's aim is to fill these research gaps by developing safety performance functions to investigate the relationship between crash frequency and their contributing factors at TAZs level, using data from Naples, Italy. In this way, the paper provides Italian local and state transportation agencies with tools to conduct proactive road safety planning.

The models were developed using recorded crashes in the period of 2009–2011. To analyze different aspects of road safety, 17 dependent variables were investigated, which were divided into six main categories: (1) crash severity; (2) vehicle type; (3) crash location; (4) crash type; (5) traffic conditions; and (6) lighting conditions. There are 53 explanatory variables, which were chosen according to previous analysis of the literature, including factors describing traffic intensity, land use, employment type, socioeconomic and demographic, and traffic network characteristics.

2. Data

The study data are relative to the city of Naples, regional capital of Campania, in Southern Italy. Naples is the third-largest municipality in Italy, with an area equal to 117,27 km^2, with 960,000 inhabitants, and a very high density—equal to 8157.79 inhabitants per km^2.

2.1. Traffic Analysis Zones

The TAZ levels adopted in the study are obtained from the layer of the 4343 census zones of the city of Naples. The first TAZ level includes 831 zones, obtained using the following zoning criteria [44,45]:

(1)   Homogeneous socioeconomic characteristics for each zone's population.
(2)   Minimizing the number of intrazonal trips.
(3)   Recognizing physical, political, and historical boundaries.
(4)   Generating only connected zones and avoiding zones that are completely contained within another zone.
(5)   Devising a zonal system in which the number of households, population, area, or trips generated and attracted are nearly equal in each zone.
(6)   Basing zonal boundaries on census zones.

Previous studies have shown that this aggregation has two disadvantages: small size in urban areas and the high percentage of zonal boundary crashes. As seen, one zoning criteria for TAZ is to minimize the number of intra-zonal trips which results in a small area size for each TAZ. Thus, it is difficult to analyze traffic crashes within these small zones at the macroscopic level. Moreover, the small size of zones creates many zones with zero crash frequencies, especially with regards to rarely occurring crashes such as severe, fatal or pedestrian crashes. The second issue is connected to zoning criteria where TAZs are often delineated by arterial roads, and therefore many crashes occur on these boundaries. The existence of boundary crashes may invalidate the assumptions of modelling which is only based on the characteristics of a zone where the crash occurred [45–47].

A simple way to overcome these two issues was proposed by Lee et al. [47] and applied in this study, which consists of aggregating contiguous small areas with similar attribute values (in our case crash characteristics). However, to meet the needs of the various dependent variables analyzed in this study, four different levels of aggregation were performed: 831, 402, 208 and 107 TAZs (see Figure 1).
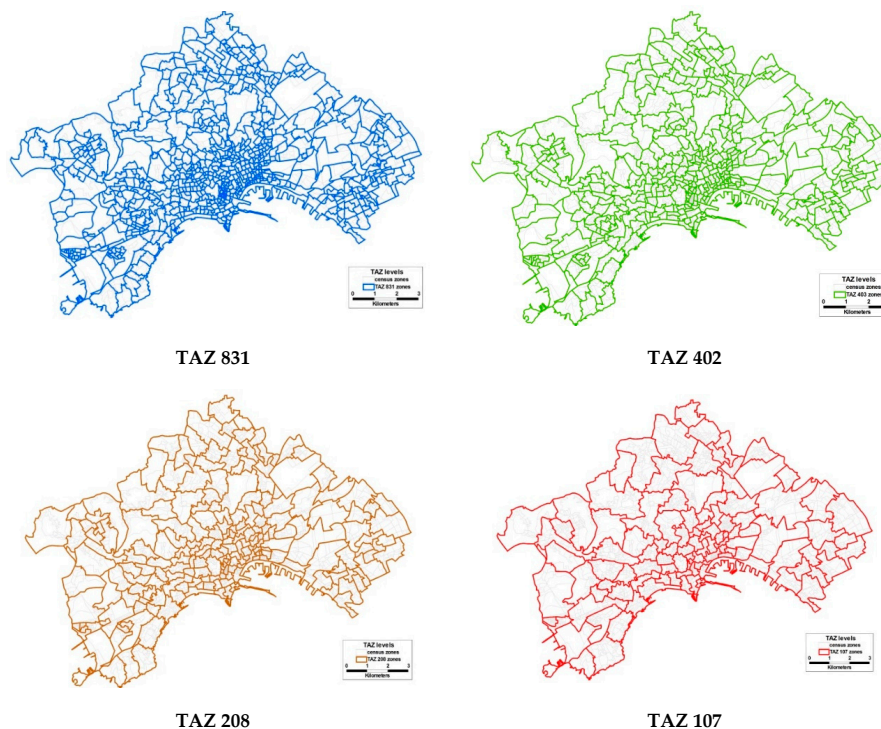

Figure 1. TAZ levels: 831, 402, 208 and 107.

2.2. Crash Data

The crash variables data were obtained from micro-data collected by the various police forces in the urban area, relative to the 3-year period from 2009 to 2011. These data include 50 fields for each crash, containing crash-related, road-related, traffic unit-related, and person-related information [48,49].

The original crash database consisted of 15,254 crashes. In order to link crashes with each TAZ, crashes needed to be geocoded on the GIS road map. However, the location was only carried out for 14,781 crashes due to missing and the poor quality data. The poor quality of information related to crash location is undoubtedly one of the most critical problems of the databases. In Italy, there are two fundamental issues related to crash location. The first concerns the incongruence between national database format and the crash report form. The Italian Database format requires the highway name, linear referencing and GPS coordinates, but the highway police crash form does not contain fields for geographical coordinates, and most of the police units do not have GPS devices. The second issue concerns missing information recorded by the different police forces. Montella et al. [49–52] found that in 36% of the crashes the location was completely missing. Due to these problems, the information reported in the database is often incomplete, making the location of accidents in some cases difficult to determine, in others impossible.

In order to provide more complete information that help to identify regions with potential safety concerns, it was decided to analyze 17 dependent variables. The dependent variables were divided into six main categories: (1) crash severity; (2) vehicle type; (3) crash location; (4) crash type; (5) traffic conditions; and (6) lighting conditions (Table 1).

Several studies have also emphasized how factors such as the exploratory variables have different impacts for each levels of severity [53]. Huang et al. [54] investigated the relations between crash frequency with a variety of aggregate road features, traffic patterns, demographic and socio-economic characteristics. In the study, models for all crash frequency and severe crash frequency were developed, and they were statistically different. Similar results were obtained by Hadayeghi et al. [7], who differentiated the dependent variables in the number of all collisions and number of fatal and injury collisions. Yasmin and Eluru [55] provide a review of earlier studies examining macro-level SPF at the various levels of injury severity. In the present study, the severity is based on the most severe injury to any person involved in the crash and was classified in three variables [50]:

— Total crashes (C);
— Property damage only (PDO);
— Severe (fatal and non-fatal) injury crashes (Cs).

Pedestrians and powered two-wheelers (PTWs) are often referred to as "vulnerable road users" [56]. The European Commission has proposed halving the overall number of road deaths in the European Union by 2020, by defining the protection of vulnerable road users as a specific objective of the road safety action program [7,55–58]. The vehicle type was distinguished into:

— Crashes where at least one car was involved (C_{car});
— Crashes where at least one truck was involved (C_{truck});
— Crashes where at least one powered two-wheeler was involved (C_{ptw});
— Crashes where at least one pedestrian was involved (C_{ped})

In literature, there are several models to estimate crash frequencies on roadway segments or at intersections [2,59–62]. However, most of these models examined the traffic safety at the microscopic level to find out factors affecting traffic crash occurrence from geometric designs and/or traffic characteristics of roadway entities, and suggested specific engineering solutions to reduce traffic crashes. To extend these analyzes at the macro level, the crash location was distinguished in:

— Crash occur on curve or tangent elements (C_{seg});

— Crash occur within intersection (C_{int}).

Previous studies have compared single-vehicle crashes with multi-vehicle crashes, and found substantial differences between these two types of crashes [63,64]. Single vehicle crashes have been shown to differ from multi-vehicle crashes in a number of aspects, which relate to road conditions, time aspects, or driver characteristics. Single-vehicle crashes are frequently associated with a disproportionate number of serious and fatal crashes. To understand the dynamics involved, the crash type was distinguished in:

— Single vehicle is a type of road traffic crash in which only one vehicle is involved (C_{sv});
— Multi-vehicle is a road traffic collision involving more than one vehicle (C_{mv}).

The literature review on the relationship between traffic volume and safety at road sections shows that crash frequency increases with increasing congestion levels [65,66]. The impact of traffic levels in urban areas during morning time period and afternoon time period on injury severity may potentially be different [67]. To explore the relationship between safety and congestion in greater detail, traffic conditions category was divided into four dependent variables:

— Crash peak day occur in the part of the day during which traffic congestion on roads is highest (from 7 a.m. to 10 a.m.) ($C_{peakday}$);
— Crashes peak night occur in the part of the night during which traffic congestion on roads is highest (from 4 p.m. to 9 p.m.) ($C_{peaknight}$);
— Crash off-peak day occur in the part of the day during which traffic congestion on roads is lower (from 10 a.m. to 4 p.m.) ($C_{off\text{-}peak\text{-}day}$);
— Crash off-peak night occur in the part of the night during which traffic congestion on roads is lower (from 9 p.m. to 4 a.m.) ($C_{off\text{-}peak\text{-}night}$).

Similarly to the previous category, the lighting conditions category studies how driver behaviour can vary throughout the day. Thus, two variables were considered:

— Total crashes during day (C_{day});
— Total crashes during night (C_{night}).

In Table 1, descriptive statistics of crash data at the TAZ level are reported. They include the total crash numbers for each dependent variable with the relative mean and standard deviation for each TAZ level.

Table 1. Descriptive statistics of crash data at the TAZ level.

			Total	TAZ831		TAZ401		TAZ208		TAZ107	
				Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
1	Crashes severity	C	14,781	17.79	26.80	36.86	47.08	71.06	72.13	138.14	117.50
2		PDO	7189	8.65	14.54	17.93	26.01	34.56	39.97	67.19	61.82
3		Cs	7592	9.14	13.29	18.93	22.64	36.50	34.55	70.95	59.32
4	Vehicle type	C_{car}	11,209	13.49	20.56	27.95	36.09	53.89	55.60	104.76	90.09
5		C_{truck}	1177	1.42	3.53	2.94	5.94	5.66	8.77	11.00	13.07
6		C_{ptw}	6332	7.62	12.25	15.79	20.91	30.44	31.85	59.18	53.67
7		C_{ped}	1768	2.13	3.95	4.41	6.62	8.50	10.30	16.52	17.73
8	Crash location	C_{seg}	8611	10.36	14.83	21.47	26.22	41.40	40.53	80.48	67.24
9		C_{int}	6170	7.42	13.97	15.39	23.60	29.66	35.80	57.66	55.29
10	Crash type	C_{sv}	6384	7.68	12.10	15.92	21.83	30.69	33.40	59.66	51.97
11		C_{mv}	8361	10.06	15.88	20.85	27.17	40.20	41.69	78.14	69.88
12	Traffic conditions	$C_{peakday}$	2002	2.41	4.03	4.99	6.64	9.63	10.46	18.71	16.53
13		$C_{peaknight}$	4126	4.97	7.69	10.29	13.33	19.84	20.25	38.56	33.12
14		$C_{off\text{-}peak\ day}$	5350	6.44	9.76	13.34	16.98	25.72	25.99	50.00	42.42
15		$C_{off\text{-}peak\ night}$	3303	3.97	6.69	8.24	11.85	15.88	17.90	30.87	28.65
16	Lighting conditions	C_{day}	9858	11.86	17.96	24.58	31.28	47.39	48.22	92.13	78.58
17		C_{night}	4923	5.92	9.48	12.28	16.63	23.67	25.24	46.01	40.64

2.3. The Explanatory Variables

The explanatory variables used in this study were carefully selected based on previous literature and their expected influence on traffic. They (see Table 2; Table 3) were then divided into three themes: Socio-Demographic (S-D), Transportation Demand Management (TDM), and Exposure.

Table 2. Explanatory variables: socio-demographic, exposure and transportation demand management.

	Variable Description	Code	Unit of Measure
Socio-Demographic			
	Children in foster care	ChFoCa	N
	Children entrusted to social care services waiting for a foster family placement	MinCollProv	N
	Children legitimized by a single parent	MinRicUnGen	N
	Children and young people reported by the Judicial Authority to the Social Services Office	MinUSSM	N
	Children included in individual tutoring projects	MinTutor	N
	Children and young people included in socio-educational projects	MinRe-edu	N
	Population	Pop	N × 1000 people
	Male population	MaPop	N × 1000 people
	Female population	FePop	N × 1000 people
	Population aged 14 and under	$0 \leq \text{Pop} < 15$	N × 1000 people
	Population aged 15 to 24	$15 \leq \text{Pop} < 25$	N × 1000 people
	Population aged 25 to 44	$25 \leq \text{Pop} < 45$	N × 1000 people
	Population aged 45 to 64	$45 \leq \text{Pop} < 65$	N × 1000 people
	Population aged 65 and above	$\text{Pop} \geq 65$	N × 1000 people
	Number of buildings	Nbuild	N
	Number of buildings for residential use	Nresbuil	N
	Surface of total housing	Ahouse	N
	Total number of dwellings	Ndwelling	N
	Number of apartments	Napartm	N
	Number of illiterate people	PopIllit	N
	Number of literature people	PopLiter	N
	Number of elementary school students	PopElem	N
	Number of secondary school students	PopSecond	N
	Number of high school students	PopHigh	N
	Number of college or graduate students	PopUniversity	N
	Average family size	FS	N
	Foreign residents in Italy–Europe	FEu	N ×1000 people
	Foreign residents in Italy–Africa	FAf	N ×1000 people
	Foreign residents in Italy–America	FAm	N ×1000 people
	Foreign residents in Italy–Asia	FAS	N ×1000 people
	Foreign residents in Italy–Oceania	FO	N ×1000 people
	Stateless persons resident in Italy	St	N ×1000 people
	Foreign residents in Italy	Fo	N ×1000 people
	Resident working in the TAZ	WKG	N ×1000 people
	Unemployed population	UNEMP	N ×1000 people
	Total number of households	Household	N ×1000 people
	Number of students	St	N
	Workforce	Wf	N
	Retired from work	RfW	N
	Population that commutes daily in the municipality of residence	PinMun	N
	Population that commutes daily outside the municipality of residence	PoutMun	N
	Workers per residents (WKG/Pop)	WKGD	-
	Population density	POPD	N/km^2
	Housing density	NDH	N/km^2
	Total neighbourhood area	AR	km^2
Exposure			
	Length of the road network	TRKM	Km
	Total vehicle kilometers traveled	VKT	veic×km
	Average congestion level	V/C	-
	Average speed	SPD	km/h
Transportation Demand Management			
	Number of trips originating in the TAZ	TRIPp	N
	Number of trips ending in the TAZ	TRIPa	N
	Number of total trips	TRIPt	N
	Number of bus stops served per hour	Hbus	N

Table 3. Descriptive statistics of explanatory variables at the TAZ level.

	TAZ831		TAZ401		TAZ208		TAZ107	
Variable	**Mean**	**St. dev.**	**Mean**	**St. dev.**	**Mean**	**St. dev.**	**Mean**	**St. dev.**
Socio-Demographic								
ChFoCa	5.09	5.63	5.33	5.63	5.49	5.48	5.25	5.28
MinCollProv	2.99	2.63	2.95	2.50	2.92	2.45	3.06	2.38
MinRicUnGen	50.55	33.31	52.68	33.16	53.71	31.87	51.04	31.36
MinUSSM	19.25	13.87	19.85	13.53	20.42	13.19	18.70	12.54
MinTutor	7.57	4.27	7.66	4.05	7.48	3.84	7.63	3.51
MinRe-edu	0.28	0.17	0.40	0.24	0.42	0.23	0.39	0.22
Pop	1.21	0.92	2.50	1.69	4.71	3.16	9.38	6.63
MaPop	0.58	0.44	1.20	0.82	2.26	1.53	4.49	3.21
FePop	0.63	0.48	1.31	0.88	2.45	1.64	4.89	3.43
$0 \leq$ Pop < 15	0.21	0.17	0.43	0.31	0.82	0.58	1.60	1.26
$15 \leq$ Pop < 25	0.17	0.14	0.35	0.25	0.66	0.49	1.30	1.05
$25 \leq$ Pop < 45	0.36	0.28	0.75	0.51	1.41	0.94	2.80	1.98
$45 \leq$ Pop < 65	0.28	0.22	0.59	0.41	1.10	0.76	2.21	1.53
Pop ≥ 65	0.19	0.16	0.39	0.30	0.71	0.54	1.46	1.05
Nbuild	1.97	4.48	4.09	8.11	8.23	13.59	15.32	21.55
Nresbuil	41.09	33.18	85.14	59.88	164.88	101.70	319.08	204.49
Ahouse	34,437.35	27,259.53	71,365.19	52,342.40	130,994.30	96,868.46	267,452.72	184,926.06
Ndwelling	435.28	319.32	902.05	596.35	1678.54	1078.21	3380.57	2183.81
Napartm	471.34	337.53	976.76	630.43	1817.06	1134.90	3660.57	2277.77
PopIllit	0.02	0.02	0.04	0.04	0.08	0.07	0.15	0.14
PopLiter	0.11	0.09	0.23	0.17	0.44	0.31	0.84	0.70
PopElem	0.27	0.24	0.57	0.43	1.11	0.81	2.13	1.80
PopSecond	0.33	0.27	0.68	0.48	1.32	0.92	2.56	2.02
PopHigh	0.28	0.26	0.57	0.51	1.03	0.93	2.14	1.71
PopUniv	0.12	0.16	0.25	0.33	0.42	0.57	0.95	1.02
FS	1201.71	910.48	2490.32	1680.33	4686.06	3149.68	9332.88	6616.97
FEu	4.07	28.67	8.44	41.24	16.20	59.84	31.64	82.43
FAf	1.66	3.53	3.43	5.40	6.81	8.44	12.86	13.79
FAm	1.78	3.14	3.68	5.27	7.01	8.65	13.80	13.50
FAS	2.99	6.88	6.20	11.59	12.01	18.82	23.24	32.68
FO	0.03	0.18	0.05	0.27	0.11	0.37	0.20	0.50
St	0.00	0.03	0.00	0.05	0.01	0.07	0.01	0.10
Fo	10.53	31.65	21.81	46.58	42.13	68.82	81.75	98.17
WKG	427.11	338.00	885.10	629.17	1650.12	1170.30	3317.07	2361.69
UNEMP	59.96	58.08	124.26	100.04	240.65	179.49	465.70	392.63
Household	224.46	172.87	465.15	312.84	886.08	579.68	1743.21	1244.77
St	84.47	71.23	175.04	137.16	319.96	254.04	655.99	479.41
Wf	218.71	183.89	453.25	352.14	829.86	643.47	1698.62	1239.80
RfW	113.96	103.66	236.16	203.57	423.93	372.60	885.05	695.41
PinMun	421.05	335.02	872.55	629.07	1617.49	1164.20	3270.04	2332.58
PoutMun	42.89	42.56	88.89	77.35	167.45	142.49	333.13	257.41
WKGD	0.35	0.06	0.35	0.05	0.35	0.04	0.35	0.04
POPD	23,191.83	19,486.86	19,885.46	16,655.95	17,689.76	15,506.82	16.06	13.10
NDH	8789.29	7518.23	7528.93	6476.56	6638.73	5996.97	6138.62	5130.71
AR	0.14	0.26	0.29	0.40	0.58	0.62	1.10	1.04
Exposure								
TRKM	1.16	1.29	2.41	2.10	4.71	3.49	9.03	6.45
VKT	1002.79	2349.33	2078.10	3690.10	4014.14	5667.34	7788.03	10,475.34
V/C	0.30	0.28	0.35	0.28	0.39	0.26	0.43	0.26
SPD	28.13	11.41	29.70	10.65	31.13	10.19	32.31	9.56
Transportation Demand Management								
TRIPp	0.07	0.06	0.14	0.10	0.25	0.17	0.51	0.34
TRIPa	0.08	0.22	0.17	0.33	0.32	0.48	0.63	0.72
TRIPt	0.15	0.24	0.31	0.36	0.58	0.53	1.14	0.85
Hbus	32.26	59.51	66.86	90.46	128.75	134.69	250.57	227.05

In Table 2, each explanatory variable is associated with their relative code and unit of measure.

In Table 3, the mean and the standard deviation of each explanatory variable are reported for each TAZ level.

The socio-demographic data were obtained from the Italian National Institute of Statistics (ISTAT) database. On a municipal level, the field of observation is made up of the habitually domiciled population (residents) as well as the population currently present [7]. The following units are measured: families; cohabitants; persons temporarily present on the census date; domiciles; other types of lodging; and buildings. For each particle census, 269 socio-demographic variables were provided. All variables consisted of measured data and most of these variables were aggregated manually to each TAZ. From a preliminary analysis of variables and a comparison with the literature, only 50 variables were used.

A further challenge to the implementation of the database is posed by the calculation of all traffic-related explanatory variables. In fact, differently from motorway contexts, which are largely monitored and have relevant traffic data readily (e.g., entry/exit counts and, mainstream sensors), urban networks require the implementation of proper models. For this reason, a state-of-the-art transport model [44,68] capable of simulating the whole multimodal transport system in the metropolitan area of Naples has been applied.

Notably, the corresponding road transport system is classified as urban, but with a significant number of motorway connections between town centre and the districts in the suburbs. At a glance, the road supply model accounts for about 54,000 nodes and 115,500 links (of which 10.91% are motorway connections and ramps), implemented in TransCAD on the basis of a TeleAtlas graph; the transit supply model is complex as well, with about 14,000 bus stops and rail stations, and 1785 road and rail transit services. The demand model—following the typical four-stage structure—has been estimated on the basis of 5000 telephone surveys and on a prior O-D matrix based on ISTAT data available on systematic trips. An extensive validation was performed using traffic counts collected at various points of the network, selected also on the basis of the methodology proposed by Simonelli et al. [69].

Overall, the morning peak hour was effectively modelled through a Stochastic User Equilibrium (SUE) assignment. This allowed for the calculation of a wide range of standard traffic-related variables for each traffic zone related to all available modes, to be included as possible explanatory variables. Finally, it is worth noting that the detailed database underlying the applied transport model also allowed for the estimation of zonal supply-related traffic variables for the transit system. More specifically, the following set of variables was calculated for each zone:

- Supply characteristics: length of the road network, number of transit services stops per hour.
- Demand characteristics: generated/attracted trips.
- Morning peak hour SUE: vehicles×km, average degree of congestion, average speed.

3. Model Development

3.1. Model Description

In this study the dependent variables have only non-negative integer values, and the statistical treatment differs from that of the normally distributed one, which cannot assume any real value, positive or negative, integer or fractional. Poisson or Negative Binomial (NB) regression models, instead, are better suited for defining the random, discrete, and nonnegative nature of crash occurrence [70]. One notable characteristic of crash-frequency data is that it is overdispersed, the variance exceeds the mean of the crash counts. When overdispersed data are present, estimating a common Poisson model can result in biased and inconsistent parameter estimates which in turn could lead to erroneous inferences regarding the factors that determine crash-frequencies. Following common practice [3,71], generalized linear modelling techniques were used to fit the models and a negative binomial distribution error structure was assumed.

The selected model form is as follows:

$$E(Y) = e^{a_0} \times e^{\sum_{i=1}^{n} b_i \times x_i} \tag{1}$$

where

$E(Y)$	=	predicted annual crash frequency,
a_i, b_i	=	model parameters, and
x_i	=	explanatory variables.

And the distribution of Y around $E(Y) = \mu$ is negative binomial with an expected value and variance of:

$$E(Y) = \mu \tag{2}$$

$$E(Y) = \mu + \frac{\mu^2}{\kappa} \tag{3}$$

where κ is the dispersion parameter of the negative binomial distribution. The modelling procedure is estimated iteratively from the model residuals, with the method of maximum likelihood being the most widely used. Because the variance decreases as κ increases, the value of κ can also be used to compare the goodness of the fit of various models fitted to the same data, in that the larger the value of κ, the smaller the variance and the better the model [72,73].

Separate multivariate models were developed for each crash variable and for each TAZ level. The model parameters and the dispersion parameter were estimated by forward stepwise selection of variables for logistic regression using an estimation of maximum likelihood [74].

The forward stepwise approach to choosing a model begins with a null model, and adds terms sequentially until further additions do not improve the fit. At each stage it selects the term which gives the greatest improvement in fit [75,76]. The decision on whether or not to keep a variable in the model was based on two criteria. The first is whether the t-ratio of the variable's estimated coefficient is significant at the 5% level. The second criterion is based in the improvement of the goodness of fit measures of the model which include that variable. A stepwise variation of this procedure retests, at each stage, terms added at previous stages to see if they are still significant. This method requires finding the value of the coefficient that maximizes the conditional likelihood.

The models were developed using the GLM (General Linear Model) procedure in SPSS software [77]. Multiplicative interaction terms were incorporated in the models, in addition to the analysis of main-effects for each variable selected. The interaction terms were created by combining two explanatory variables at a time and considering all the possible combinations.

3.2. Measuring Goodness of Fit

Several measures can be used to assess the goodness of fit of the models [78,79]. To measure the goodness of fit in linear regression models the coefficient of determination R^2 is often used. However, the R^2 measure is only appropriate to linear regression, with its continuous dependent variables. The R^2 statistic is a measure of the percentage of unconditional variance of the dependent variable explained by the available covariates. It is considered meaningful only in measuring the goodness of fit of normal linear regression models with additive mean functions, in which the conditional variance of the dependent variable is not a function of its conditional mean. Because crash prediction models are nonnormal and functional forms are typically nonlinear, the R^2 is not appropriate as a goodness of fit measure. To get around this problem, a number of statisticians have developed other goodness of fit measures, such as: Pseudo R^2, normalized adjusted R^2 (R_n^2), dispersion parameter-based R^2 (R_α^2), Akaike information criterion (AIC) and Bayesian information criterion (BIC). In this study, it was chosen to use the R_α^2 and AIC, being the goodness of fit measures most commonly used in this analysis type [80–82].

The R_α^2 uses the size of dispersion parameter in the conventional negative binomial regression model as a yardstick to determine how well the variance of the data is explained, which is calculated as follows [82]:

$$R_\alpha^2 = 1 - \left(\frac{\kappa_{min}}{\kappa} \right) \tag{4}$$

where:

κ_{min} = smallest dispersion parameter possible that is obtained by having no covariates in the model (by assuming that all sites have an identical prediction estimate equal to the mean over all sites) and

κ = dispersion parameter for the calibrated model.

For a given data set, the largest dispersion parameter value is first estimated by fitting the observed data Y, with a negative binomial distribution (which includes no covariate). The main advantage of this measure is its simplicity in addition to being bound between 0 when no covariate is included, and 1 when covariates are perfectly specified.

The AIC is an estimator of the relative quality of statistical models, the smaller the statistic the better the model [83,84]. The AIC value is calculated as follows:

$$AIC = -2 \times ML + 2 \times p \qquad (5)$$

ML = maximum log-likelihood of the fitted model and

p = number of parameters in the model.

The first term in the AIC equation measures the badness of fit, or bias, when the maximum likelihood estimates of the parameters are used. The second term measures the complexity of the model, thus penalizing the model for using more parameters. The goal for selecting the best model is to choose the best fit with the least complexity.

4. Results and Discussion

The models were developed using the stepwise forward procedure, adding one explanatory variable at each step. 68 regression models were developed in order to examine the relationships between zonal crashes and a suite of factors describing traffic intensity, land use, employment type, socioeconomic and demographic, and traffic network characteristics.

Results of the stepwise procedure for all crash variables are shown in the tables below. Table 4 shows the results of the crash severity group: total crashes (C); property damage only (PDO); severe injury crashes (C_s). Table 5 shows the results of the crash vehicle type group: crashes where at least one car was involved (C_{car}); crashes where at least one truck was involved (C_{truck}); crashes where at least one powered two wheeler was involved (C_{ptw}); crashes where at least one pedestrian was involved (C_{ped}). Table 6 shows the results of the crash location group: crash occur on curve or tangent elements (C_{seg}); crash occur within intersection (C_{int}). Table 7 shows the results of the vehicle type group are reported: single vehicle is a type of road traffic crash in which only one vehicle is involved (C_{sv}); multi-vehicle is a road traffic collision involving more than one vehicle (C_{mv}). Table 8 shows the results of the traffic conditions group are reported: peak day crashes occur in the part of the day during which traffic congestion on roads is highest (from 7 a.m. to 10 a.m.) ($C_{peakday}$); peak night crashes occur in the part of the night during which traffic congestion on roads is highest (from 4 p.m. to 9 p.m.) ($C_{peaknight}$); off-peak day crashes occur in the part of the day during which traffic congestion on roads is lower (from 10 a.m. to 4 p.m.) ($C_{off\text{-}peak\text{-}day}$); off-peak night crashes occur in the part of the night during which traffic congestion on roads is lower (from 9 p.m. to 4 a.m.) ($C_{off\text{-}peak\text{-}night}$). Table 9 shows the results of the crash lighting condition group: crashes during the day (C_{day}); crashes during the night (C_{night}).

Table 4. SPFs (crash severity group): Parameter estimates and good of fit measures.

	C				PDO				Cs			
Variable	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**
Intercept	0.62	1.7	2.22	3.15	−0.11	0.95	1.44	2.36	−0.48	1.13	1.7	2.51
	(0.12)	(0.15)	(0.19)	(0.23)	(0.13)	(0.16)	(0.2)	(0.23)	(0.25)	(0.16)	(0.19)	(0.21)
MinRe-edu	−0.34	−0.58	−0.57	−0.52	−0.59	−0.87	−0.83	−0.7	−0.1	−0.46	−0.34	
	(0.21)	(0.17)	(0.21)	(0.26)	(0.22)	(0.02)	(0.22)	(0.28)	(0.22)	(0.18)	(0.21)	
Pop ≥ 65	1.27	0.34	0.29	0.09	1.24	0.34	0.29		1.05	0.26	0.31	0.13
	(0.23)	(0.08)	(0.1)	(0.06)	(0.24)	(0.09)	(0.1)		(0.27)	(0.1)	(0.1)	(0.06)
TRKM	0.23	0.11	0.05	0.02	0.27	0.13	0.07	0.03	0.19	0.07	0.03	
	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.02)	
V/C	0.91	0.51	0.51	0.5	0.93	0.48	0.48	0.34	0.86	0.55	0.52	0.46
	(0.13)	(0.14)	(0.18)	(0.23)	(0.14)	(0.15)	(0.19)	(0.23)	(0.14)	(0.15)	(0.18)	(0.22)
SPD	0.03	0.03	0.02	0.02	0.03	27.9	0.03	0.03	0.03	0.02	0.02	0.02
	(0)	(0)	(0.01)	(0.01)	(0)	(4.52)	(0.01)	(0.01)	(0)	(0)	(0.01)	(0.01)
TRIPp									1.21	1.15		
									(0.85)	(0.57)		
TRIPa								0.18	0.44		0.21	0.18
								(0.1)	(0.33)		(0.12)	(0.09)
TRIPt	0.74	0.35	0.21		0	0.31	0.21					
	(0.32)	(0.18)	(0.11)		(0)	(0.17)	(0.11)					
Hbus	7.36	5.57	3.49	1.97	0.01	5.11	3.31	1.82	7.43	6.08	3.48	1.97
	(1.03)	(0.71)	(0.54)	(0.38)	(0)	(0.73)	(0.55)	(0.38)	(1.03)	(0.7)	(0.54)	(0.35)
1/k	0.78	0.49	0.35	0.27	0.82	0.51	0.36	0.28	0.78	0.51	0.35	0.25
	(0.06)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
R_a^2	0.50	0.58	0.61	0.61	0.54	0.62	0.65	0.64	0.51	0.55	0.57	0.61
AIC	5834	3349	1997	1169	4673	2770	1693	1016	4878	2890	1751	1030

Note: Standard errors of the parameter estimates are reported in brackets.

Table 5. SPFs (vehicle type group): Parameter estimates and good of fit measures.

	C_{car}				C_{truck}				C_{ptw}				C_{ped}			
Variable	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**
Inter	−0.41	0.45	1.59	2.65	−2.75	−2.13	−2.55	−0.07	−0.02	1	1.35	2.5	−0.74	0.17	0.72	1.57
	(0.24)	(0.3)	(0.17)	(0.21)	(0.38)	(0.46)	(0.68)	(0.26)	(0.14)	(0.17)	(0.19)	(0.26)	(0.17)	(0.21)	(0.27)	(0.26)
MinRe-edu									−0.71	−0.83	−0.59	−0.49	−0.59	−0.61	−0.43	−0.77
									(0.24)	(0.2)	(0.22)	(0.29)	(0.31)	(0.26)	(0.32)	(0.34)
MaPop									0.32	0.13	0.06					
									(0.09)	(0.05)	(0.04)					
Pop ≥ 65	1.31	0.37	0.31										1.08	0.32	0.39	0.28
	(0.23)	(0.09)	(0.1)										(0.31)	(0.12)	(0.13)	(0.08)
WKGD	1.72	0.13			2.24	2.22	4.79									
	(0.59)	(0.02)			(0.94)	(1.16)	(1.62)									
TRKM	0.28	0.13	0.06		0.24	0.12	0.08		0.12	0.04	0.04		−0.14	−0.14	−0.09	
	(0.04)	(0.02)	(0.02)		(0.05)	(0.03)	(0.02)		(0.04)	(0.03)	(0.02)		(0.04)	(0.03)	(0.02)	
V/C	0.84	0.57	0.53	0.32	1.17	1.09	1.4	1.45	1.04	0.71	0.56	0.48	0.94	0.79	0.88	1.18
	(0.13)	(0.14)	(0.17)	(0.21)	(0.18)	(0.21)	(0.24)	(0.27)	(0.15)	(0.17)	(0.19)	(0.25)	(0.19)	(0.21)	(0.26)	(0.33)
SPD	0.03	0.03	0.02	0.02	0.03	0.02	0.03	25.24	26.23	0.02	0.02	0.02	0.02	0.02	0.01	
	(0)	(0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(7.7)	(4)	(0)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	
TRIPa			0.21										1.18	0.74	0.48	
			(0.09)										(0.51)	(0.32)	(0.22)	
TRIPt	0.49	0.29	0.23		0.62	0.68	0.2	0.21	1.02	0.46	0.29	0.18				
	(0.31)	(0.17)	(0.11)		(0.39)	(0.29)	(0.1)	(0.1)	(0.37)	(0.23)	(0.12)	(0.1)				
Hbus	6.5	5.2	3.44	1.92	5.69	4.67	3.06	0.03	0.01	6.96	3.42	2.38	9.61	7.33	4.26	1.59
	(1.04)	(0.73)	(0.53)	(0.37)	(1.29)	(0.96)	(0.63)	(0.01)	(0)	(0.78)	(0.54)	(0.41)	(1.37)	(1.01)	(0.75)	(0.42)
1/k	0.81	0.51	0.35	0.27	1.04	0.69	0.43	0.31	0.87	0.57	0.34	0.33	1.18	0.83	0.59	0.45
	(0.05)	(0.04)	(0.04)	(0.04)	(0.11)	(0.09)	(0.07)	(0.06)	(0.06)	(0.05)	(0.04)	(0.05)	(0.1)	(0.09)	(0.08)	(0.07)
R_a^2	0.52	0.58	0.63	0.62	0.62	0.66	0.69	0.69	0.51	0.56	0.66	0.56	0.42	0.48	0.43	0.49
AIC	5388	3138	1877	1107	2236	1492	996	645	4546	2735	1596	1007	2921	1878	1228	764

Note: Standard errors of the parameter estimates are reported in brackets.

Table 6. SPFs (crash location group): Parameter estimates and good of fit measures.

	C_{seg}				C_{int}			
Variable	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**
Intercept	−0.9	0.1	−0.52	−0.08	−0.35	0.26	0.96	1.91
	(0.24)	(0.3)	(0.55)	(0.67)	(0.15)	(0.18)	(0.22)	(0.24)
Pop					0.12	0.07	0.03	
					(0.05)	(0.03)	(0.02)	
Pop ≥ 65	1.43	0.26		0.21				
	(0.23)	(0.11)		(0.09)				
WKGD	2.92	2.7	6.1	7.2				
	(0.6)	(0.74)	(1.38)	(1.63)				
TRKM	0.28	0.1	0.07	0.05	0.2	0.07		
	(0.04)	(0.02)	(0.02)	(0.01)	(0.04)	(0.03)		
V/C	0.75	0.58	0.59	0.5	1.1	0.84	0.86	0.78
	(0.14)	(0.16)	(0.2)	(0.22)	(0.17)	(0.17)	(0.21)	(0.25)
SPD	0.03	0.02	0.02	0.02	0.03	0.03	0.03	0.03
	(0)	(0)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	(0.01)
TRIPp	1.64	1.21	0.72	0.01				
	(0.9)	(0.58)	(0.36)	(0)				
TRIPa					1.02	0.46		0.21
					(0.51)	(0.26)		(0.12)
Hbus	7.67	5.8	3.1	0	7.13	6.16	5.14	2.37
	(1.05)	(0.75)	(0.59)	(0)	(1.25)	(0.86)	(0.58)	(0.38)
1/k	0.82	0.58	0.41	0.24	1.21	1.54	0.49	0.37
	(0.05)	(0.05)	(0.04)	(0.04)	(0.08)	(0.72)	(0.05)	(0.05)
R_α^2	0.49	0.52	0.55	0.63	0.46	0.53	0.56	0.57
AIC	5042	2994	1813	1051	4465	2703	1662	997

Note: Standard errors of the parameter estimates are reported in brackets.

Table 7. SPFs (vehicle type group): Parameter estimates and good of fit measures.

	C_{sv}				C_{mv}			
Variable	**831**	**404**	**208**	**107**	**831**	**404**	**208**	**107**
Intercept	−0.53	1.13	1.03	2.38	−0.09	−0.06	1.49	2.33
	(0.27)	(0.15)	(0.61)	(0.2)	(0.13)	(0.31)	(0.18)	(0.23)
MinRe-edu	−0.28	−0.79	−0.01					
	(0.23)	(0.19)	(0)					
Pop ≥ 65	0.78	0.18			1.51	0.46		
	(0.23)	(0.09)			(0.25)	(0.09)		
WKGD	1.41		2.28			2.46		
	(0.64)		(1.43)			(0.8)		
TRKM	0.28	0.13	0.08	0.02	0.19	0.08	0.03	
	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.02)	
V/C	0.83	0.41	0.41		0.96	0.72	0.54	0.5
	(0.14)	(0.15)	(0.2)		(0.15)	(0.15)	(0.19)	(0.23)
SPD	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.02
	(0)	(0)	(0.01)	(0.01)	(0)	(0)	(0.01)	(0.01)
TRIPa			0.24					
			(0.1)					
TRIPt	0.46	0.26			0.73	0.37	0.43	0.24
	(0.3)	(0.17)			(0.35)	(0.19)	(0.14)	(0.1)
Hbus	7.01	5.22	0	1.79	7.07	5.56	4.09	2.25
	(1.04)	(0.73)	(0)	(0.39)	(1.11)	(0.77)	(0.6)	(0.39)
1/k	0.76	0.51	0.39	0.67	0.92	0.56	0.41	0.31
	(0.05)	(0.04)	(0.04)	(0.29)	(0.06)	(0.05)	(0.04)	(0.05)
R_α^2	0.53	0.58	0.58	0.57	0.49	0.58	0.58	0.59
AIC	4561	2734	1685	1000	4970	2926	1783	1055

Note: Standard errors of the parameter estimates are reported in brackets.

Table 8. SPFs (traffic conditions group): Parameter estimates and good of fit measures.

Variable	$C_{peakday}$ 831	404	208	107	$C_{peaknight}$ 831	404	208	107	$C_{off-peak\ day}$ 831	404	208	107	$C_{off-peak\ night}$ 831	404	208	107
Intercept	−1.06 (0.16)	−0.13 (0.18)	0.22 (0.21)	1.3 (0.23)	−1.32 (0.25)	−0.66 (0.31)	−0.92 (0.54)	−0.41 (0.71)	−0.2 (0.13)	0.79 (0.16)	1.24 (0.19)	2.01 (0.21)	−0.56 (0.15)	0.39 (0.17)	1 (0.21)	1.57 (0.28)
MinRe-edu	−0.76 (0.26)	0 (0)		−0.67 (0.28)					−0.39 (0.22)	−0.62 (0.18)	−0.47 (0.21)		−0.58 (0.26)	0 (0)	−0.87 (0.24)	
Pop	0.1 (0.06)	0.09 (0.03)	0.03 (0.02)												0.04 (0.02)	
Pop ≥ 65									1.04 (0.27)	0.29 (0.1)	0.3 (0.1)	0.12 (0.06)				
MaPop*25 ≤ Pop < 45													0.28 (0.08)	0.07 (0.03)		
WKGD					2.44 (0.61)	2.68 (0.75)	5.17 (1.36)	5.68 (1.75)								
TRKM	0.2 (0.04)	0.1 (0.02)	0.04 (0.01)	0.02 (0.01)	0.21 (0.04)	0.09 (0.02)	0.06 (0.02)	0.03 (0.01)	0.21 (0.04)	0.1 (0.02)	0.04 (0.02)		0.25 (0.04)	0.11 (0.03)	0.05 (0.02)	0.06 (0.01)
V/C	0.92 (0.16)	0.65 (0.17)	0.72 (0.22)	0.58 (0.23)	0.79 (0.14)	0.02 (0)	0.6 (0.18)	0.56 (0.22)	0.89 (0.14)	0.56 (0.15)	0.62 (0.18)	0.6 (0.23)	0.82 (0.16)	0.42 (0.18)	0.33 (0.21)	
SPD	0.03 (0)	0.02 (0)	0.02 (0.01)	0.02 (0.01)	0.03 (0)	0.02 (0)	0.02 (0.01)	0.03 (0.01)	0.03 (0)	0.02 (0)	0.02 (0.01)	0.02 (0.01)	0.03 (0)	0.03 (0.01)	0.03 (0.01)	0.04 (0.01)
TRIPp	1.49 (1)				2.03 (0.7)	1.22 (0.47)			0 (0)	0.93 (0.56)						
TRIPa												0.22 (0.09)				
TRIPt							0.2 (0.12)	0.14 (0.09)			0.25 (0.11)					
Hbus	6.32 (1.07)	5.2 (0.73)	3.78 (0.56)	1.88 (0.34)	7.92 (1.01)	5.86 (0.7)	3.19 (0.56)	1.55 (0.36)	0.01 (0)	5.6 (0.69)	3.25 (0.52)	1.92 (0.36)	8.16 (1.22)	6.1 (0.79)	3.82 (0.57)	
1/k	0.78 (0.07)	0.48 (0.06)	0.37 (0.05)	0.24 (0.04)	0.73 (0.05)	0.48 (0.06)	0.34 (0.04)	0.23 (0.04)	0.74 (0.05)	0.48 (0.04)	0.32 (0.04)	0.26 (0.04)	1 (0.07)	0.61 (0.06)	0.41 (0.05)	0.5 (0.07)
R_{α}^2	0.56	0.63	0.62	0.65	0.55	0.61	0.63	0.63	0.54	0.60	0.65	0.63	0.50	0.57	0.60	0.35
AIC	2993	1903	1231	754	3941	1903	1492	904	4322	2593	1589	955	3646	2258	1427	909

Note: Standard errors of the parameter estimates are reported in brackets.

Table 9. SPFs (lighting conditions group): Parameter estimates and good of fit measures.

Variable	C_{day} 831	404	208	107	C_{night} 831	404	208	107
Intercept	0.35 (0.13)	1.25 (0.16)	1.88 (0.19)	2.57 (0.21)	−0.2 (0.14)	0.76 (0.16)	1.97 (0.2)	2.09 (0.22)
MinRe-edu	−0.48 (0.21)	−0.77 (0.18)	−0.68 (0.21)		−0.48 (0.24)	−0.83 (0.2)	−0.78 (0.21)	
Pop	0.12 (0.05)	0.1 (0.02)	0.05 (0.02)					
25 ≤ Pop < 45							0.13 (0.06)	
MaPop*25 ≤ Pop < 45					0.25 (0.08)	0.06 (0.03)		
TRKM	0.18 (0.04)	0.09 (0.02)	0 (0)		0.25 (0.04)	0.11 (0.03)	0.05 (0.02)	0.02 (0.01)
V/C	0.9 (0.13)	0.62 (0.14)	0.56 (0.18)	0.53 (0.21)	0.77 (0.15)	0.39 (0.17)	0.59 (0.19)	
SPD	0.03 (0)	0.03 (0)	0.02 (0.01)	0.02 (0.01)	0.03 (0)	0.03 (0)	0.02 (0.01)	0.02 (0.01)
TRIPp	1.65 (0.84)							
TRIPa	0.66 (0.36)		0.25 (0.12)					
TRIPt		0.31 (0.17)		0.22 (0.09)				
Hbus	7.15 (1.01)	5.48 (0.7)	3.51 (0.54)	2.12 (0.35)	8.61 (1.14)	6.29 (0.76)	3.87 (0.53)	2.4 (0.38)
1/k	0.76 (0.05)	0.48 (0.04)	0.37 (0.04)	0.26 (0.04)	0.91 (0.06)	0.58 (0.05)	0.36 (0.04)	0.32 (0.05)
R_{α}^2	0.52	0.59	0.62	0.62	0.51	0.57	0.63	0.56
AIC	5232	3042	1835	1078	4198	2548	1591	952

Note: Standard errors of the parameter estimates are reported in brackets.

Analysis of the results shows that the goodness of fit of the models improves with decreasing the number of TAZ, particularly R_α^2 increases and AIC decreases, except for C_{ptw}, $C_{off\text{-}peak\ day}$, and $C_{off\text{-}peak\ night}$, where going from 208 to 107 TAZ R_α^2 decreases. Observing the parameters of good fit, the best TAZ is 208. This finding is consistent with previous studies. Xu et al. [85,86] observed that zoning schemes with the higher number of zones tend to have an increasing number of significant variables, more stable coefficient estimation, smaller standard error, but worse model performance. Moreover, Lee et al. [47] confirmed that a higher level of aggregation of TAZ provides the best estimation models with less dispersion, but also demonstrated that if the zone is too large it may lose many local features.

For all models, exposure variables were the most significant predictors and positively associated with the number of crashes in each TAZ, as suggested and frequently shown in the literature [9]. In all models, exposure variables such as the length of the road network (TRKM), the average congestion level (V/C) and the average speed (SPD) all gave significant results. These outcomes are consistent with the literature. Lovegrove et al. [87] and Wei and Lovegrove [88] found that regional congestion levels (V/C) were directly associated with the crash prediction model, and estimated that decreasing V/C values would result in decreasing crash estimates. This suggested that the average weighted V/C value for a given traffic zone could be used as a surrogate indicator of road safety. Xie et al. [89] found that street length has a positive impact on crash occurrence.

In the model, the statistically significant demographic variables are resident population (POP), population aged 65 and above (Pop $\geq$ 65), male population (MaPop), and population aged 25 to 45 (25 $\leq$ Pop < 45). In particular, Pop $\geq$ 65 is associated with eight dependent variables: C, PDO, C_s, C_{car}, C_{ped}, C_{seg}, C_{sv} and $C_{off\text{-}peak\ day}$. These results are in line with other studies such as Montella et al. [57], in which the older population showed greater propensity toward fatal crashes. The studies conducted by Noland and Quddus [19] and Aguero-Valverde and Jovanis [12] showed that a higher percentage of the elderly population are associated with a higher number of road crashes, while, according to Amoh-Gyimah et al. [9] the elderly population percentage was positively associated with minor injury pedestrian crashes. A possible explanation is that the elderly may have weak eyesight and might usually take longer to cross a street, thus increasing their exposure to vehicle traffic [90]. $C_{peakday}$ is associated with resident population (POP), which was consistent with the research of Abdel-Aty et al. [10,91], Hadayeghi et al. [21], and Xie et al. [89]. The male population (MaPop) variable affects CPTW. Montella et al. [56] found that male PTW drivers, in combination with other variables, was significantly correlated with fatal crashes. Interaction male population and population aged 25 to 44 (MaPop*25 $\leq$ Pop < 45) variable affects $C_{off\text{-}peak\ night}$ and C_{night}. These outcomes are consistent with many studies, which have shown that as you get older, aggressive driving tendencies decrease and driver gender is correlated directly with aggressive driving [92,93]

The results also showed that increased crashes were associated with increases in workers per residents (WKGD). In particular, workers per residents (WKGD) is associated with C_{car}, C_{truck}, C_{seg}, $C_{peaknight}$ and C_{day}. These results confirm earlier research by Lovegrove et al. [15] and Kim et al. [39,94]. The difference between $C_{peaknight}$ and $C_{peakday}$ is related to the different activities carried out during the hours of the day. During the peak night hours, work-related trips prevail, while during the hours of the morning trips are more diverse, such as travel to school or shops.

The children and young people included in socio-educational projects (MinRe-edu) variable negatively affects the frequency of crashes. In particular MinRe-edu is associated with C, PDO, C_s, C_{ptw}, C_{ped}, $C_{peakday}$, $C_{off\text{-}peak\ day}$, $C_{off\text{-}peak\ night}$, C_{day} and C_{night}. These projects provide care to children from disadvantaged neighbourhoods, and organize after school educational activities and workshops which include music, art, cooking, sports, games and leisure activities. This variable shows that these projects also promote less aggressive driving habits in young people. The results of the present study confirmed the positive effects of an active learning-based educational program [95]. This result highlight that the road user is the first link in the road safety chain. Whatever the technical measures

in place, the effectiveness of a road safety policy depends ultimately on the users' behaviour. For this reason, education, training and enforcement are essential [58].

Regarding the transportation network, it was observed that as the number of trips increases, crashes also tend to increase. In particular, total trips ($TRIP_t$) is associated with C, PDO, C_{car}, C_{truck}, C_{ptw}, C_{sv}, C_{mv} and C_{day}. $TRIP_p$ is associated with C_{seg}, $TRIP_a$, C_s, C_{ped} and C_{int}. These results confirm earlier research by Abdel-Aty et al. [91]; Dong et al. [96], Naderan and Shahi [97]. A certain TDM scenario may be developed to reduce trips of a specific purpose, and the related number of crashes could be predicted. Hbus is the number of bus stops served in one hour in the area and it is an indirect measure of bus stop capacity. Hbus is a proxy for pedestrian traffic, so the positive sign indicates a growing correlation with crashes. The association between increased collisions and increased bus stops (BS) is consistent with researches of Kim et al. [39,94], Wei et al. [88] and Rhee et al. [98]. A larger number of subway stations were found to increase traffic crashes. Bus stops attract pedestrian activities, and an increase of such nodes would increase the possibility of conflict between pedestrians and vehicle traffic, and at bus stops, between buses, other vehicles and pedestrians [89]. Pedestrian traffic is most likely unprotected, and therefore pedestrian routes must be improved.

5. Conclusions

Incorporating safety considerations into the transportation planning process in a comprehensive way has emerged as a strategy for improving transportation safety in recent years.

Macro-level safety performance functions were developed in this study to provide decision support tools for planners to consider safety in the transportation planning process, and to promote more sustainable land use and transport patterns. The objective of this study was to develop a series of macro-level safety performance functions that are consistent with aggregate travel demand models. It might be very helpful for administrations which do not have quality crash data to identify the area which has the highest number of crashes.

68 models were developed using recorded crashes in the period 2009–2011 in the city of Naples. To analyze different aspects of road safety, 17 dependent variables were investigated for four TAZ levels. The first result obtained highlights that, observing parameters of good fit of 68 models, the optimal scale was the TAZ with 208 zones. This result shows that using traditional zoning schemes might not be the optimal systems for regional safety analysis. In this study, the optimal zoning was obtained by aggregating contiguous small areas with similar crash characteristics.

The main significant variables were: children and young people included in socio-educational projects, population, population aged 65 and above, population aged 25 to 44, male population, total vehicle kilometers traveled, average congestion level, average speed, number of trips originating in the TAZ, number of trips ending in the TAZ, number of total trips and, number of bus stops served per hour.

Most of these variables are consistent with the literature, except for the MinRe-edu variable (children and young people included in socio-educational projects). Although a large number of road safety education programs exist, very few studies use crashes as an evaluation criterion—most use intermediate variables such as knowledge, attitudes and (self-reported) safe behaviour [95,99]. This study highlights the positive influence of socio-educational projects, connecting the presence of such projects to a reduction in crashes.

The findings of this study highlight that road safety management must take a more comprehensive approach with broader range of policy tools that can be applied to a wider range of component parts that comprise the road system. This approach must recognize that the road user is the first link in the road safety chain, and to achieve greater safety, socio-educational projects have to be included. This is in line with the first objective of the Road Safety Action Program 2011–2020 proposed by European Commission: Improve education and training of road users [58].

References

1. Hadayeghi, A.H.; Shalaby, A.S.; Persaud, B.N. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accid. Anal. Prev.* **2010**, *42*, 676–688. [CrossRef]
2. American Association of State Highway and Transportation Officials (AASHTO). *Highway Safety Manual*, 1st ed.; AASHTO: Washington, DC, USA, 2010; Volumes 1–3.
3. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. [CrossRef]
4. Saha, D.; Alluri, P.; Gan, A.; Wu, W. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. *Accid. Anal. Prev.* **2018**, *118*, 166–177. [CrossRef] [PubMed]
5. Lee, J.; Abdel-Aty, M.; Jiang, X. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accid. Anal. Prev.* **2015**, *78*, 146–154. [CrossRef] [PubMed]
6. Cai, Q.; Abdel-Aty, M.; Lee, J. Comparative analysis of zonal systems for macro-level crash modeling. *J. Saf. Res.* **2017**, *61*, 157–166. [CrossRef] [PubMed]
7. Hadayeghi, A.H.; Shalaby, A.S.; Persaud, B.N. Safety prediction models: A proactive tool for safety evaluation in urban transportation planning applications. *Transp. Res. Rec.* **2007**, *2019*, 225–236. [CrossRef]
8. Wedagama, D.P.; Bird, R.N.; Metcalfe, A.V. The influence of urban land-use on non-motorised transport casualties. *Accid. Anal. Prev.* **2006**, *38*, 1049–1057. [CrossRef]
9. Amoh-Gyimah, R.; Saberi, M.; Sarvi, M. Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accid. Anal. Prev.* **2016**, *93*, 147–159. [CrossRef]
10. Abdel-Aty, M.; Lee, J.; Siddiqui, C.; Choi, K. Geographical unit based analysis in the context of transportation safety planning. *Transp. Res. Part A Policy Pract.* **2013**, *49*, 62–75. [CrossRef]
11. Aguero-Valverde, J. Multivariate spatial models of crash frequency at area level: Case of Costa Rica. *Accid. Anal. Prev.* **2013**, *59*, 365–373. [CrossRef]
12. Aguero-Valverde, J.; Jovanis, P.P. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accid. Anal. Prev.* **2006**, *38*, 618–625. [CrossRef]
13. Li, Z.; Wang, W.; Liu, P.; Bigham, J.M.; Ragland, D.R. Using geographically weighted Poisson regression for county-level crash modeling in California. *Saf. Sci.* **2013**, *58*, 89–97. [CrossRef]
14. Siddiqui, C.; Abdel-Aty, M.; Huang, H. Aggregate nonparametric safety analysis of traffic zones. *Accid. Anal. Prev.* **2012**, *45*, 317–325. [CrossRef]
15. Lovegrove, G.R.; Sayed, T. Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Can. J. Civ. Eng.* **2006**, *33*, 609–621. [CrossRef]
16. Lovegrove, G.R.; Sun, J. Using community-based macro-level collision prediction models to evaluate the safety level of neighbourhood road network patterns. In Proceedings of the 89th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 10–14 January 2010; pp. 1–17.
17. Chen, H.Y.; Senserrick, T.; Martiniuk, A.L.C.; Ivers, R.Q.; Boufous, S.; Chang, H.Y.; Norton, R. Fatal crash trends for Australian young drivers 1997–2007: Geographic and socioeconomic differentials. *J. Saf. Res.* **2010**, *41*, 123–128. [CrossRef]
18. Pirdavani, A.; Daniels, S.; Van Vlierden, K.; Brijs, K.; Kochan, B. Socioeconomic and sociodemographic inequalities and their association with road traffic injuries. *J. Transp. Health* **2017**, *4*, 152–161. [CrossRef]
19. Noland, R.B.; Quddus, M.A. A spatially disaggregate analysis of road casualties in England. *Accid. Anal. Prev.* **2004**, *36*, 973–984. [CrossRef]
20. Ladron de Guevara, F.; Washington, S.P.; Oh, J. Forecasting crashes at the planning level simultaneous negative binomial crash model applied in Tucson, Arizona. *Transp. Res. Rec.* **2004**, *1897*, 191–199. [CrossRef]
21. Hadayeghi, A.; Shalaby, A.S.; Persaud, B.N. Macrolevel accident prediction models for evaluating safety of urban transportation systems. *Transp. Res. Rec.* **2003**, *1840*, 87–95. [CrossRef]

22. Gomes, M.M.; Pirdavani, A.; Brijs, T.; Pitombo, C.S. Assessing the impacts of enriched information on crash prediction performance. *Accid. Anal. Prev.* **2019**, *122*, 162–171. [CrossRef]

23. Pulugurtha, S.S.; Duddu, V.R.; Kotagiri, Y. Traffic analysis zone level crash estimation models based on land use characteristics. *Accid. Anal. Prev.* **2013**, *50*, 678–687. [CrossRef]

24. Lee, J.; Abdel-Aty, M.; Choi, K.; Huang, H. Multi-level hot zone identification for pedestrian safety. *Accid. Anal. Prev.* **2015**, *76*, 64–73. [CrossRef]

25. Siddiqui, C.; Abdel-Aty, M.; Choi, K. A comparison of geographical unit-based macro-level safety modeling. In Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, USA, 22–26 January 2012; pp. 1–18.

26. Wier, M.; Weintraub, J.; Humphreys, E.H.; Seto, E.; Bhatia, R. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accid. Anal. Prev.* **2009**, *41*, 137–145. [CrossRef]

27. Ukkusuri, S.; Hasan, S.; Aziz, H.A. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. *Transp. Res. Rec.* **2011**, *2237*, 98–106. [CrossRef]

28. Lascala, E.A.; Gerber, D.; Gruenewald, P.J. Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis. *Accid. Anal. Prev.* **2000**, *32*, 651–658. [CrossRef]

29. Wang, X.; Jin, Y.; Abdel-Aty, M.; Tremont, P.J.; Chen, X. Macro level model development for safety assessment of road network structures. *Transp. Res. Rec.* **2012**, *2280*, 100–109. [CrossRef]

30. Pirdavani, A.; Brijis, T.; Bellemans, T.; Kochan, B.; Wets, G. Application of different exposure measures in development of planning-level zonal crash prediction models. *Transp. Res. Rec.* **2012**, *2280*, 145–153. [CrossRef]

31. Pirdavani, A.; Brijs, T.; Bellemans, T.; Kochan, B.; Wets, G. Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. *Accid. Anal. Prev.* **2013**, *50*, 186–195. [CrossRef]

32. Karim, A.; Wahba, M.M.; Sayed, T. Spatial effect on the zone level collision prediction models. *Transp. Res. Rec.* **2014**, *2398*, 50–59. [CrossRef]

33. Huang, H.; Xu, P.; Abdel-Aty, M. Transportation safety planning: A spatial analysis approach. In Proceedings of the 92nd Annual Meeting of the Transportation Research Board, Washington, DC, USA, 13–17 January 2013; pp. 1–20.

34. Kmet, L.; Brasher, P.; Macarthur, C. A small area study of motor vehicle fatalities in Alberta, Canada. *Accid. Anal. Prev.* **2003**, *35*, 177–182. [CrossRef]

35. Washington, S.; Schalwyk, I.V.; Meyer, M.; Dumbaugh, E.; Zoll, M. *NCHRP Report 546: Incorporating Safety into Long-Range Transportation Planning*; Transportation Research Board of the National Academies: Washington, DC, USA, 2006. [CrossRef]

36. Park, S.H.; Kim, D.-K.; Kho, S.-Y.; Rhee, S. Identifying hazardous locations based on severity scores of highway crashes. In Proceedings of the 12th World Conference on Transport Research (WCTR), Lisbon, Portugal, 11–15 July 2010.

37. Wei, V.F.; Lovegrove, G. Sustainable road safety: A new (?) neighbourhood road pattern that saves VRU lives. *Accid. Anal. Prev.* **2012**, *44*, 140–148. [CrossRef]

38. Sun, J. Sustainable Road Safety: Development, Transference and Application of Community-Based Macro-Level Collision Prediction Models. Master's Thesis, University of British Columbia, Vancouver, BC, Canada, May 2009. [CrossRef]

39. Kim, K.; Made Brunner, I.; Yamashita, E.Y. Influence of land use, population, employment, and economic activity on accidents. *Transp. Res. Rec.* **2006**, *1953*, 56–64. [CrossRef]

40. Quddus, M.A. Modeling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accid. Anal. Prev.* **2008**, *40*, 1486–1497. [CrossRef]

41. Kim, K.; Yamashita, E. Motor Vehicle Crashes and Land Use. Empirical Analysis from Hawaii. *Transp. Res. Rec.* **2002**, *1784*, 72–79. [CrossRef]

42. Liu, J.; Khattak, A.J.; Wali, B. Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. *Accid. Anal. Prev.* **2017**, *109*, 132–142. [CrossRef]

43. Shirazi, M.; Lord, D.; Geedipally, S.R. Sample-size guidelines for recalibrating crash prediction models: Recommendations for the Highway Safety Manual. *Accid. Anal. Prev.* **2016**, *93*, 160–168. [CrossRef]

44. Cascetta, E. *Transportation Systems Analysis: Models and Application*, 2nd ed.; Springer: New York, NY, USA, 2009; ISBN 978-0-387-75856-5.

45. Lee, J.; Abdel-Aty, M.; Jiang, X. Development of zone system for macro-level traffic safety analysis. *J. Transp. Geogr.* **2014**, *38*, 13–21. [CrossRef]

46. Xu, P.; Huang, H.; Dong, N.; Abdel-Aty, M. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accid. Anal. Prev.* **2014**, *70*, 110–120. [CrossRef]

47. Lee, J.; Abdel-Aty, M.; Cai, Q. Intersection crash prediction modeling with macro-level data from various geographic units. *Accid. Anal. Prev.* **2017**, *102*, 213–226. [CrossRef] [PubMed]

48. National Institute of Statistics (Istat). Population and Households. Available online: https://www.istat.it/en/population-and-households (accessed on 27 March 2019).

49. Montella, A.; Andreassen, D.; Tarko, A.; Turner, S.; Mauriello, F.; Imbriani, L.; Romero, M.; Singh, R. Crash databases in Australasia, the European union, and the United States. *Transp. Res. Rec.* **2013**, *2386*, 128–136. [CrossRef]

50. Montella, A.; Andreassen, D.; Tarko, A.; Turner, S.; Mauriello, F.; Imbriani, L.; Romero, M.; Singh, R. Critical review of the international crash databases and proposals for improvement of the Italian national database. *Procedia-Soc. Behav. Sci.* **2012**, *53*, 49–61. [CrossRef]

51. Montella, A.; Chiaradonna, S.; Criscuolo, G.; De Martino, S. Development and evaluation of a web-based software for crash data collection, processing and analysis. *Accid. Anal. Prev.* **2017**, in press. [CrossRef] [PubMed]

52. Montella, A.; Chiaradonna, S.; Criscuolo, G.; De Martino, S. Perspectives of a web-based software to improve crash data quality and reliability in Italy. In Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017; pp. 451–456. Available online: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8005714 (accessed on 27 March 2019). [CrossRef]

53. Xie, B.; An, Z.; Zheng, Y.; Li, Z. Incorporating transportation safety into land use planning: Pre-assessment of land use conversion effects on severe crashes in urban China. *Appl. Geogr.* **2019**, *103*, 1–11. [CrossRef]

54. Huang, H.; Abdel-Aty, M.; Darwiche, A. County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transp. Res. Rec.* **2010**, *2148*, 27–37. [CrossRef]

55. Yasmin, S.; Eluru, N. Latent segmentation based count models: Analysis of bicycle safety in Montreal and Toronto. *Accid. Anal. Prev.* **2016**, *95*, 157–171. [CrossRef] [PubMed]

56. Montella, A.; Aria, M.; D'Ambrosio, A.; Mauriello, F. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accid. Anal. Prev.* **2012**, *49*, 58–72. [CrossRef]

57. Montella, A.; Aria, M.; D'Ambrosio, A.; Mauriello, F. Data-mining techniques for exploratory analysis of pedestrian crashes. *Transp. Res. Rec.* **2011**, *2237*, 107–116. [CrossRef]

58. European Commission. *Towards a European Road Safety Area: Policy Orientations on Road Safety 2011–2020*; Commission Communication (2010) 398 Final; European Commission: Brussels, Belgium, 20 July 2010.

59. Le, T.; Gross, F.B.; Persaud, B.; Eccles, K.A.; Soika, J. *Safety Evaluation of Multiple Strategies at Signalized Intersections (No. FHWA-HRT-17-062)*; Federal Highway Administration: McLean, VA, USA, 2018.

60. Montella, A.; Imbriani, L.L. Safety performance functions incorporating design consistency variables. *Accid. Anal. Prev.* **2015**, *74*, 133–144. [CrossRef]

61. Montella, A.; Colantuoni, L.; Lamberti, R. Crash prediction models for rural motorways. *Transp. Res. Rec.* **2008**, *2083*, 180–189. [CrossRef]

62. Persaud, B.; Nguyen, T. Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. *Transp. Res. Rec.* **2014**, *1635*, 113–120. [CrossRef]

63. Wu, Q.; Zhang, G.; Zhu, X.; Liu, X.C.; Tarefder, R. Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accid. Anal. Prev.* **2016**, *94*, 35–45. [CrossRef]

64. Ulfarsson, G.F.; Mannering, F.L. Differences in male and female injury severities in sport-utility vehicle minivan, pickup and passenger car accidents. *Accid. Anal. Prev.* **2004**, *36*, 135–147. [CrossRef]

65. Cafiso, S.; Cava, G.; Montella, A. Safety inspections as supporting tool for safety management of low-volume roads. *Transp. Res. Rec.* **2011**, *2203*, 116–125. [CrossRef]

66. Chang, G.L.; Xiang, H. *The Relationship between Congestion Levels and Accidents*; Research Report No. MD-03-SP 208B46; Maryland State Highway Administration: Baltimore, MD, USA, 2003.

67. Pahukula, J.; Hernandez, S.; Unnikrishnan, A. A time of day analysis of crashes involving large trucks in urban areas. *Accid. Anal. Prev.* **2015**, *75*, 155–163. [CrossRef] [PubMed]

68. Ortuzar, J.; Willumsen, L. *Modelling Transport*, 4th ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2011; ISBN 978-0-470-76039-0.

69. Simonelli, F.; Marzano, V.; Papola, A.; Vitiello, I. A network sensor location procedure accounting for o-d matrix estimate variability. *Transp. Res. B-Meth.* **2012**, *46*, 1624–1638. [CrossRef]

70. Milton, J.; Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* **1998**, *25*, 395–413. [CrossRef]

71. Mannering, F.L.; Bhat, C.R. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [CrossRef]

72. Hauer, E.; Ng, J.C.N.; Lovell, J. Estimation of safety at signalized intersections. *Transp. Res. Rec.* **1988**, *1185*, 48–61.

73. Hauer, E. *Observational before–after Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*, 1st ed.; Emerald Group Publishing Limited: Bingley, UK, 1997; ISBN 978-0080430539.

74. Hosmer, D.W., Jr.; Wang, C.Y.; Lin, I.C.; Lemeshow, S. A computer program for stepwise logistic regression using maximum likelihood estimation. *Comput. Programs Biomed.* **1978**, *8*, 121–134. [CrossRef]

75. Agresti, A. *Categorical Data Analysis*, 2nd ed.; John Wiley: Hoboken, NJ, USA, 2002.

76. Jobson, J.D. *Applied Multivariate Data Analysis, Volume II, Categorical and Multivariate Methods*; Springer Science & Business Media: New York, NY, USA, 2012.

77. IBM. Generalized Linear Models Model. Available online: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/advanced/idh_idd_genlin_model.html#idh_idd_genlin_model (accessed on 27 March 2019).

78. Lyon, C.; Persaud, B. Pedestrian collision prediction models for urban intersections. *Transp. Res. Rec.* **2002**, *1818*, 102–107. [CrossRef]

79. Hauer, E.; Harwood, D.W.; Council, F.M.; Griffith, M.S. Estimating safety by the empirical bayes method: A tutorial. *Transp. Res. Rec.* **2002**, *1784*, 126–131. [CrossRef]

80. Hauer, E. *The Art of Regression Modeling in Road Safety*, 2015 ed.; Springer: New York, NY, USA, 2015.

81. Montella, A.; Imbriani, L.L.; Marzano, V.; Mauriello, F. Effects on speed and safety of point-to-point speed enforcement systems: Evaluation on the urban motorway A56 Tangenziale di Napoli. *Accid. Anal. Prev.* **2015**, *75*, 164–178. [CrossRef]

82. Miaou, S.P.; Lu, A.; Lum, H.S. Pitfalls of using R^2 to evaluate goodness of fit of accident prediction models. *Transp. Res. Rec.* **1996**, *1542*, 6–13. [CrossRef]

83. Montella, A.; Persaud, B.; D'Apuzzo, M.; Imbriani, L. Safety evaluation of automated section speed enforcement system. *Transp. Res. Rec.* **2012**, *2281*, 16–25. [CrossRef]

84. Pardillo Mayora, J.M.; Manzo, R.B.; Orive, A.C. Refinement of accident prediction models for Spanish national network. *Transp. Res. Rec.* **2006**, *1950*, 65–72. [CrossRef]

85. Xu, P.; Huang, H.; Dong, N. The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research. *J. Traffic Transp. Eng.* **2018**, *5*, 73–82. [CrossRef]

86. Xu, P.; Huang, H.; Dong, N.; Wong, S. Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach. *Accid. Anal. Prev.* **2017**, *98*, 330–337. [CrossRef] [PubMed]

87. Lovegrove, G.R.; Lim, C.; Sayed, T. Using Macro-level collision prediction models to conduct a road safety evaluation of a regional transportation plan. In Proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 13–17 January 2008; pp. 1–23.

88. Wei, F.; Lovegrove, G. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. *Accid. Anal. Prev.* **2013**, *61*, 129–137. [CrossRef] [PubMed]

89. Xie, K.; Ozbay, K.; Yang, H. A multivariate spatial approach to model crash counts by injury severity. *Accid. Anal. Prev.* **2019**, *122*, 189–198. [CrossRef] [PubMed]

90. Demetriades, D.; Murray, J.; Martin, M.; Velmahos, G.; Salim, A.; Alo, K.; Rhee, P. Pedestrians injured by automobiles: Relationship of age to injury type and severity. *J. Am. Coll. Surg.* **2004**, *199*, 382–387. [CrossRef] [PubMed]

91. Abdel-Aty, M.; Siddiqui, C.; Huang, H. Integrating trip and roadway characteristics in managing safety at traffic analysis zones. *Transp. Res. Rec.* **2011**, *2213*, 20–28. [CrossRef]

92. Fuller, R. Towards a general theory of driver behavior. *Accid. Anal. Prev.* **2005**, *37*, 461–472. [CrossRef] [PubMed]

93. Agerwala, S.M.; Votta, A.; Hogan, B.; Yannocone, J.; Samuels, S.; Chiffriller, S. Aggressive driving in young motorists. *Int. J. Hum. Soc. Sci. Res.* **2008**, *2*, 182–185. [CrossRef]

94. Kim, K.; Pant, P.; Yamashita, E. Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii. *Transp. Res. Rec.* **2010**, *2147*, 9–17. [CrossRef]

95. Zare, H.; Niknami, S.; Heidarnia, A.; Fallah, M.H. Traffic safety education for child pedestrians: A randomized controlled trial with active learning approach to develop street-crossing behaviors. *Transp. Res. F-Traf.* **2019**, *60*, 734–742. [CrossRef]

96. Dong, N.; Huang, H.; Xu, P.; Ding, Z.; Wang, D. Evaluating spatial-proximity structures in crash prediction models at the level of traffic analysis zones. *Transp. Res. Rec.* **2014**, *2432*, 46–52. [CrossRef]

97. Naderan, A.; Shah, J. Aggregate crash prediction models: Introducing crash generation concept. *Accid. Anal. Prev.* **2010**, *42*, 339–346. [CrossRef]

98. Rhee, K.A.; Kim, J.; Lee, Y.; Ulfarsson, G.F. Spatial regression analysis of traffic crashes in Seoul. *Accid. Anal. Prev.* **2016**, *91*, 190–199. [CrossRef]

99. Dragutinovic, N.; Twisk, D. *The Effectiveness of Road Safety Education: A Literature Review*; Report No R-2006-6; SWOV Institute for Road Safety Research: The Hague, The Netherlands, 2006; pp. 1–85.