

Article

# Investigating the Risk Factors Associated with the Severity of the Pedestrians Injured on Spanish Crosstown Roads

Natalia Casado-Sanz <sup>1,\*</sup> , Begoña Guirao <sup>1</sup> , Antonio Lara Galera <sup>2</sup> and Maria Attard <sup>3</sup> 

<sup>1</sup> Departamento de Ingeniería del Transporte, Territorio y Urbanismo, ETSI Caminos, Canales y Puertos, Universidad Politécnica de Madrid, 28040 Madrid, Spain; begona.guirao@upm.es

<sup>2</sup> Departamento de Ingeniería Civil: Construcción, ETSI Caminos, Canales y Puertos, Universidad Politécnica de Madrid, 28040 Madrid, Spain; antoniolorenzo.lara@upm.es

<sup>3</sup> Institute for Climate Change and Sustainable Development, OH132, University of Malta, MSD2080 Msida, Malta; maria.attard@um.edu.mt

\* Correspondence: natalia.casado.sanz@upm.es; Tel.: +34-913-366-784

Received: 8 August 2019; Accepted: 18 September 2019; Published: 22 September 2019



**Abstract:** According to the Spanish General Traffic Accident Directorate, in 2017 a total of 351 pedestrians were killed, and 14,322 pedestrians were injured in motor vehicle crashes in Spain. However, very few studies have been conducted in order to analyse the main factors that contribute to pedestrian injury severity. This study analyses the accidents that involve a single vehicle and a single pedestrian on Spanish crosstown roads from 2006 to 2016 (1535 crashes). The factors that explain these accidents include infractions committed by the pedestrian and the driver, crash profiles, and infrastructure characteristics. As a preliminary tool for the segmentation of 1535 pedestrian crashes, a *k*-means cluster analysis was applied. In addition, multinomial logit (MNL) models were used for analysing crash data, where possible outcomes were fatalities and severe and minor injured pedestrians. According to the results of these models, the risk factors associated with pedestrian injury severity are as follows: visibility restricted by weather conditions or glare, infractions committed by the pedestrian (such as not using crossings, crossing unlawfully, or walking on the road), infractions committed by the driver (such as distracted driving and not respecting a light or a crossing), and finally, speed infractions committed by drivers (such as inadequate speed). This study proposes the specific safety countermeasures that in turn will improve overall road safety in this particular type of road.

**Keywords:** pedestrian safety; contributing factors; crash severity; multinomial logit; crosstown roads

## 1. Introduction

There has been an increase in the number of road traffic accidents worldwide, making road safety a great concern. According to the World Health Organization, the number of annual road traffic deaths reached 1.35 million in 2018, which is considered to be the eighth leading cause of death globally [1]. Pedestrians, cyclists, and motorcyclists disproportionately suffer most of these accidents, which accounts for more than half of global road traffic fatalities and hence they are considered vulnerable users by most traffic administrations.

Pedestrians are considered the most fragile road users in the transport system. They are at maximum risk compared to any other road users because of their fragility, slow pace, and their absence of protection [2]. In Europe, the safety of a pedestrian has been problematic for a long time. The actions taken to reduce pedestrian crashes have been much less notable compared to those for the total traffic accidents, although the total number of fatalities has decreased significantly during the period

2006–2016. In the European Union, a total of 5320 pedestrians were killed in road accidents in 2016, 21% of all road fatalities [3].

One of the European countries that has been giving great importance to pedestrian safety in the past few years is Spain. Pedestrians contribute to half of all deaths (51%) in Spanish urban areas, which is the second-highest percentage in the whole of the EU following Latvia (58%). Furthermore, in this region, one among every five traffic accident fatalities is a pedestrian. According to the Spanish General Traffic Accident Directorate, in 2017 a total of 351 pedestrians were killed and 14,322 pedestrians were injured in pedestrian-vehicle crashes in Spain [4]. One part of these accidents took place in crosstown roads, a particular type of road with a high case fatality rate. Generally, crosstown roads are defined as sections of the road network that pass through towns without an alternative to bypass the centre, thus causing conflicts between urban mobility requirements and the higher speed demanded by interurban traffic. As a result, the main street is used as an interurban rural route. This impedes the pedestrian routes and, thus, affects inhabitants' day to day life. In addition, these roads are usually located in small towns where walking is the most common mode of transport because of the absence or limitations in rural public transport services. Furthermore, it is important to note that traffic accidents not only cause fatalities or injuries but also incur considerable economic losses. In the context of traffic accidents, in the year 2010, it was estimated that the economic loss associated with each fatality was as high as 1.3 million euros in Spain, while for a seriously injured person it was 219,000 euros and for a lightly injured person it was 6,100 euros [5]. The general economic burden related to traffic accident fatalities is highly worrying even though these costs differ for each country.

Considering the above facts, it is important to identify and characterise the risk factors that contribute to pedestrian-vehicle crash injury severity. This is important to determine interventions and could support traffic engineers, planners, and decision-makers to consider the contributing factors in engineering countermeasures. Consequently, the primary objective of this paper is to identify different contributing factors that increase the probability of a fatal outcome given that a pedestrian-vehicle crash has occurred on Spanish crosstown roads. In order to achieve this objective, a cluster analysis is carried out as a preliminary tool for segmenting pedestrian crash data. In addition, multinomial logit (MNL) models were used to identify the primary factors in pedestrian crash severity.

In order to describe the research as a whole, this paper is organized as follows. Section 1 presents an introduction with the context and the objective of the study. Section 2 provides a brief literature review of the methodologies commonly used to analyse accidents. The database and the statistical techniques used for this case study are described in Section 3. Section 4 provides the results and a discussion of the analysis. Finally, Section 5 presents the main research conclusions.

## 2. Literature Review

Pedestrian crash analysis is one of the topics that has received special interest among the traffic safety researchers in recent years. Exploring the attributes of pedestrian-vehicle accidents, the characterization of the factors that contributed to the injury severity levels and the prediction of pedestrian-vehicle collisions were the most studied features. The existing research and literature demonstrate an extensive diversity of factors that contribute to the occurrence and the severity of pedestrians involved in motor vehicle accidents, such as behavioural factors, road design, and environmental conditions.

Applying an ordered probit model, in 2005, Lee and Abdel-Aty (2005) [6] investigated vehicle-pedestrian collisions at intersections in Florida for a period of three years. Moreover, their study reported that pedestrian and driver characteristics (such as older pedestrians and pedestrians under the effects of alcohol), vehicle size (larger than passenger cars), and environment conditions (such as adverse weather and dark lighting) generally worsen the injury of accidents. In order to study crash injury severity, in 2008 Kim et al. [7] investigated single-vehicle single-pedestrian collisions occurred in North Carolina in the period 1997–2000. The results of their study reveal that parameters such as the age of the pedestrians, male drivers, two-way roads, overspeed, dark-lighted condition,

and commercial areas, among others, are the main factors increasing the probability of fatal pedestrian injury. In addition, Ulfarsson et al. (2010) [8] used the same database to examine the fault allocation of pedestrian-vehicle accidents. Finally, it was concluded that drivers were responsible for their manoeuvres, and pedestrians were blamed for those cases they were distractedly crossing streets.

In 2010, Abdul-Aziz et al. [9] conducted a similar study. They analysed pedestrian-vehicle accidents in New York in the period 2002–2006 and demonstrated that roadway features (such as the number of lanes, road surface, and light condition), traffic attributes (such as the type of vehicles, signal control), and land-use features (parking facilities, commercial area, and so on) contribute to severe injuries. Similarly, another research was conducted by Pour-Rouholamin and Zhou (2016) [10]. They analysed single pedestrian-single vehicle crashes in Illinois for a period of four years. They reported that pedestrians over 65 years old, pedestrians not wearing contrasting clothing, adult drivers, summer season, time of day, multilane highways, darkness, and collisions with pickup, were factors that contributed to more severe injuries. In order to analyse the contributing causes that affect pedestrian injury severity in rural Connecticut, Ivan et al. [11] developed an ordered probit model in 2000. They also reported that vehicle type, drivers under the influence of alcohol and elderly pedestrians significantly increase pedestrian injury severity. Similarly, another study [12] reported that dark lighting conditions increase the probability of major injuries of pedestrian accidents. This study also demonstrated that crashes on two-lane roads have higher probability of no-injury in urban areas. In addition, other parameters, such as the presence of intersections without traffic lights or the absence of pedestrian crossings, were also analysed, and these factors were associated with fatal crashes [2,13]. Similarly, it was noted that the severity of pedestrian injury is also associated with vehicle type [14]. This study compiled collision data collected by the competent authority, trauma registry, and autopsy during 1995–1999 in Maryland. It was concluded that sport utility vehicles (SUV) and vans contribute significantly to injury severity as compared to other vehicle types. All these studies mentioned above are based on the factors that most contribute to the severity of the pedestrian injury, with most focusing uniquely on urban areas. As far as it is known, very few studies have looked at crosstown roads [15]. This type of road is considered a hybrid because of its urban and rural characteristics and traffic.

There is a large number of statistical techniques, such as binary logistic regression [13], ordered probit models [6,16], mixed logit models [17,18], and multinomial logit models [19,20], which can be applied to explore crash severity. However, traffic accidents often happen under different conditions, which make traffic safety data deeply heterogeneous and thus difficult to model [21]. As a result of this, data mining methods such as clustering and classification techniques have emerged and been combined with classic statistical methods. Moreover, it is not always possible to ensure that each segment consists of a homogeneous group of accidents, hence it is better to reduce heterogeneity by fragmenting the data [22].

Cluster analysis is a multivariate technique that is primarily applied to group objects that form conglomerates [23]. This technique is based on a taxonomy that maximises the similarity within in-cluster components and the dissimilarity between inter-cluster factors [24]. This statistical technique has been applied widely in the field of road safety analysis. Pardillo-Mayora et al. (2010) [25] applied cluster techniques for examining the accident rate on two-lane rural roads in Spain and investigated the effect of roadside features on safety. Their study described the main roadside attributes that affect the outcomes of roadway departures and groups these features into a ranking which exhibits uniform effects on the frequency of run-off-road accidents with injuries. The results create a five-level roadside hazardousness index, which is considered a useful tool for roadside design and the planning of safety improvements. Furthermore, cluster analysis was applied by Karlaftis and Tarko (1998) [26] for classifying 92 areas of Indiana into three different types of areas: urban, suburban, and rural. A negative binomial regression model was then applied for analysing the influence that the age of drivers had on traffic crashes. Interestingly, their results revealed considerable statistical differences between the models applied to all the data sets and the models based on clusters. However, some researchers have used other models, such as latent class cluster, which is a probability-model-based cluster analysis

method in which the class memberships can be inferred from the observed variables, more accurately. In order to identify seven clusters and analyse the severity of different types of traffic accidents, Depaire et al. (2008) [22] applied this technique in combination with multinomial logit (MNL) models and demonstrated the importance of segmenting the data in the road safety analysis. In addition, in order to analyse the main factors in pedestrian crash severity, Sun et al. (2019) [27] applied a latent class cluster model as a preliminary tool for segmenting 14,236 pedestrian crashes in Louisiana. Their results demonstrated the importance of the application of these clustering techniques, which help in identifying hidden relationships in traffic safety analyses. Finally, for analysing the seriousness of traffic accidents in Spain, De Oña et al., 2013 [28] applied latent class cluster techniques in combination with Bayesian networks. Furthermore, they applied latent class clustering as a preceding tool for fragmenting 3229 accidents on rural roads in the region of Granada in Spain for the period 2005–2008. The results demonstrated that both statistical techniques collectively provide more information as compared to the one that would have been obtained without a previous division of the data [28].

In relation to data, the level of quality and the breakdown of the database determines the statistical methods and the validity of the findings, which can help and guide the authorities in the development of strategic plans and the implementation of measures to improve road safety and thus reduce accidents. It should be noted that any method used is limited by the restrictions of the database. Nevertheless, there is a disparity concerning data uniformity between countries, and even between local jurisdictions in the same country [29]. Even several road fatalities and victim meanings have been debated, looking for a means of standardisation. For example, an international comparison of different definitions of seriously injured has been carried out by Utriainen et al. (2018) [30]. Based on the Montella et al. (2012) study [29], Table 1 has been designed to show an international comparison of the variables gathered in the main guidelines and databases of different countries (US, New Zealand, and Australian databases, together with the requirements of the EU Directive) and to position Spain in the framework of the road safety. For the purposes of this research, it is noted that the Spanish dataset does not contain information about traffic and road layout at the accident location. This can be considered a weakness in the Spanish dataset.

**Table 1.** International comparison between elements in the official road crash databases. Source: Adapted from Montella et al. (2012) [29].

Variable	EU Directive <sup>a</sup>	US MMUCC <sup>b</sup>	Australia	New Zealand	Spain
Crash location	Precise as possible location	Road name, GPS coordinates	Road name, reference point, distance, direction	Road name, GPS coordinates	Road name, km
Crash narrative	No	No	Yes	Yes	Yes
Crash sketch	No	No	Yes, access restricted	Yes	Yes
Crash type	Yes	Recorded in the traffic units section	Yes	Yes	Yes
Collision type	Yes	8 descriptors	Yes	Yes	33 descriptors
Contributing circumstances	No	Environmental circumstances	Yes	Yes	Yes
Weather conditions	Yes	10 descriptors	Yes	5 descriptors	9 descriptors
Light conditions	Yes	7 descriptors	Yes	7 descriptors	Yes
Reported crashes	Not specified	All severities	All injury severities	All severities	All severities
Definition of non-fatal injury levels	Severe and non-severe injuries	A: Suspected serious injury B: Suspected minor injury C: Possible injury	Injured, admitted to hospital Injured, required medical treatment	Serious: Requiring medical treatment Minor: other injuries	Hospitalised, injured Non-hospitalised, injured
Fatalities	Within 30 days	Within 30 days	Within 30 days	Within 30 days	Within 30 days
Link with hospital data	No	No	In Western Australia	No	Yes
Contributing circumstances	No	11 descriptors	No	Numerous cause codes	Yes
Speed limit	Yes	Yes	Yes	Yes	Yes <sup>d</sup>
Surface conditions	Yes	10 descriptors	Yes	3 descriptors	9 descriptors

Table 1. Cont.

Variable	EU Directive <sup>a</sup>	US MMUCC <sup>b</sup>	Australia	New Zealand	Spain
Road curve	No	Yes	Yes	4 descriptors	5 descriptors
Road segment gradient	No	Yes	No	No	No
Age	Yes	Date of birth	Yes	Yes <sup>c</sup>	Yes
Gender	Yes	Yes	Yes	Yes	Yes
Nationality	Yes	No	Foreign drivers identified	Foreign drivers identified	Yes
Injury status	No	5 descriptors	4 descriptors	Yes	5 descriptors
Driver action	No	19 descriptors	In crash narrative	In crash narrative	23 descriptors
Pedestrian action	No	11 descriptors	In crash narrative	In crash narrative	11 descriptors
Violation codes	No	Yes	Yes	Yes	No
Alcohol level	Yes	Yes	Yes	Yes	Yes
Drug test results	No	Yes	Yes	Yes	Yes
Safety equipment	Yes	Yes	Yes	Yes	Yes
Seating position	No	Yes	Yes	Yes	Yes
ADT <sup>e</sup>	No	Yes	No	Yes	No
Curve radius	No	Yes	No	Yes	No
Length	No	Yes	No	Yes	No

<sup>a</sup> Directive 2008/96/EC of the European Parliament and of the Council of 19 November 2008 on road infrastructure safety management, <sup>b</sup> MMUCC Guideline. Model minimum uniform crash criteria. National Highway Traffic Safety Administration (NHTSA), <sup>c</sup> only pedestrian and cyclist ages in coded crash listing. Other ages in police crash reports, <sup>d</sup> speed limit recorded since 2015, <sup>e</sup> average daily traffic.

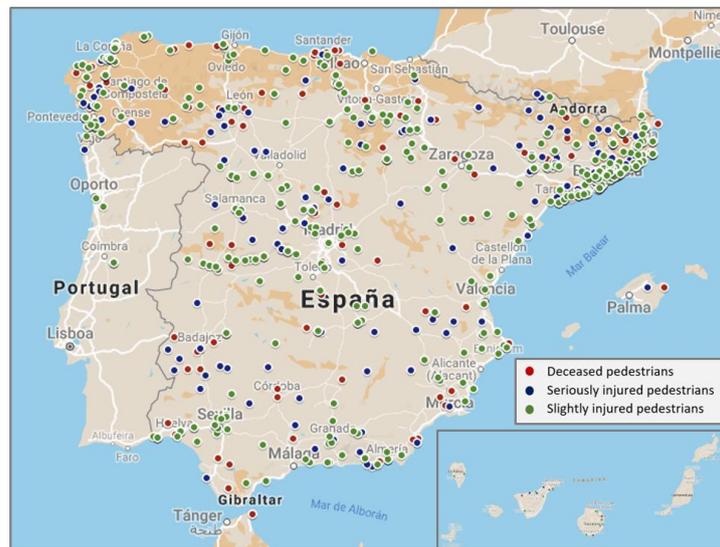
As far as it is known, the segmentation of accidents by using cluster methods and subsequent statistical analysis has never been applied to accidents on crosstown roads, making this study of vehicle-pedestrian crashes on Spanish crosstown roads a pioneering one.

### 3. Materials and Methods

Therefore, the primary objective of this work is to analyse different factors that contribute to increasing the probability of a fatal outcome in the case of a crash that involves pedestrians on Spanish crosstown roads. In addition, clustering techniques are examined for consolidating the existing results pertaining to the suitability of segmenting accident data. The data used in this research, collected from the Spanish Accident Statistics database, include accidents on Spanish crosstown roads for a period of 11 years (2006–2016), which involved a single vehicle and a single pedestrian. The selected sample consists of more than 90% of the accidents on crosstown roads, which involved pedestrians (the number of accidents with more than one pedestrian injured as well as more than one vehicle involved is very small). However, one of the disadvantages of the available database is that there is no information about the values of the various variables in cases when no accidents take place. The ideal would have been to contemplate those situations in which a pedestrian over 65 years old crosses a crosstown road without being run over. It would require the design of another more complex survey where data must be collected using other types of methodologies (for example, pedestrian tracking using mobile devices and apps). The available database only allows us to estimate the severity of an accident involving pedestrians.

In the final data set for the model estimation, a total of 1535 accidents that involved pedestrian-vehicle crashes were considered after removing the crashes with incomplete data. Each observation of the sample records the severity of the injury of each pedestrian involved in an accident along with a set of parameters that include pedestrian and driver data, vehicle characteristics and road infrastructure features. As a result, the final sample consisted of 189 accidents in which the pedestrian was killed (12.4% of the total sample), 452 crashes in which the pedestrian was seriously injured (29.4% of the total sample), and 894 accidents in which the pedestrian was slightly injured (58.2% of the total sample). Different locations of these accidents are shown in Figure 1. The pedestrian injury severity involving a single-vehicle collision, where the injury could be fatal, severe, or minor, is considered the dependent variable. An overview of the descriptive statistics of pedestrian-vehicle crashes and all variables used for this research is presented in Table 2. Similar to most of the other countries,

all of these variables (such as age and gender of victims, lane width, shoulder type, road markings, and so on) are automatically collected by the Spanish Directorate General of Traffic (DGT). It would be very interesting to add other variables, such as territorial or exposure variables, which can enrich the database under the study, but these require extensive resources since they would be collected manually. This study therefore only focuses on analysing the recorded variables.



**Figure 1.** Location and distribution of the pedestrian-vehicle crashes in Spanish crosstown roads.

### 3.1. Cluster Analysis

Clustering is a data mining technique that manages a collection of unlabelled data. The primary objective of this technique is to group the data objects into different clusters, and each cluster shows common features with the data from which it is extracted. Moreover, there are several types of clustering techniques that follow different approaches. The  $k$ -means clustering technique is applied in this study, which divides  $n$  observations into  $k$  clusters in which each element belongs to the cluster with the nearest mean. The algorithm applies the following steps. First, all points are placed in the space represented by the objects that are grouped. This set of points is termed as the centroids of the initial group. Next, each object is assigned to the group that has the closest centroid. The positions of the  $K$  centroids are recalculated once all the objects have been assigned. Finally, this process is repeated until it is confirmed that the centroids no longer move. Moreover, the distance between objects of different groups is developed. The metric that minimises this distance can be calculated.

In order to identify homogeneous groups, the software SPSS Statistics v24 was used in this study. To calculate the distance, the squared error cost function was used, which is expressed as follows [29]:

$$E = \sum_{j=1}^k \sum_{i=1}^N (\|v_i - c_j\|)^2 \quad (1)$$

where  $N$  is the number of data,  $k$  is the number of centres,  $\|v_i - c_j\|$  is a selected distance measure between a data point  $v_i$  and the cluster centre and  $c_j$  is an indicator of the distance of the  $n$  data points from their respective cluster centres. Based on MacQueen, 1967 [31], the Euclidean distances between the data sample and all the centres are calculated and the nearest centre is modified:

$$\Delta C_z(t) = \eta(t)[v(t) - c_z(t-1)] \quad (2)$$

where  $z$  indicates the nearest centre to the data  $v(t)$ . The centres and the data are expressed in terms of time  $t$ , where  $c_z(t-1)$  represents the centre location at the previous clustering step.

**Table 2.** Descriptive statistics of pedestrian-vehicle crashes and classification by severity.

Variable	No. of Crashes	Fatal Injury	Severe Injury	Minor Injury
<b>Driver's age</b>				
<18 years old	23	8.70%	30.40%	60.90%
18–30 years old	347	13.50%	37.20%	49.30%
31–64 years old	934	13.20%	27.60%	59.20%
65 and over	231	8.20%	27.70%	64.10%
<b>Pedestrian's age</b>				
<18 years old	294	1.70%	26.50%	71.80%
18–30 years old	158	4.40%	27.20%	68.40%
31–64 years old	487	10.30%	28.50%	61.20%
65 and over	596	21.60%	33.20%	45.20%
<b>Driver's gender</b>				
Male	1199	13.70%	29.60%	56.70%
Female	336	8.00%	30.70%	61.30%
<b>Pedestrian's gender</b>				
Male	767	13.30%	29.70%	57.00%
Female	768	11.60%	29.90%	58.50%
<b>Atmospheric factors</b>				
Good weather	1305	12.20%	29.70%	58.10%
Light rain	132	15.90%	28.80%	55.30%
Heavy rain	41	9.70%	41.50%	48.80%
Fog	5	0.00%	80.00%	20.00%
Snow	3	0.00%	66.70%	33.30%
Hail	1	0.00%	100.00%	0.00%
Heavy wind	12	8.30%	33.40%	58.30%
Other	36	16.70%	13.90%	69.40%
<b>Day of the week</b>				
Beginning of week (Mon)	230	12.60%	32.60%	54.80%
Weekday (Tue, Wed, Thu)	712	13.30%	28.50%	58.20%
End of week (Fri)	269	11.20%	27.90%	60.90%
Weekend (Sat, Sun)	324	11.40%	32.40%	56.20%
<b>Type of day</b>				
Holiday	210	12.40%	35.70%	51.90%
Working day	927	12.60%	28.10%	59.30%
Eve of holiday	190	10.00%	29.50%	60.50%
Day after a holiday	208	13.90%	32.20%	53.90%
<b>Lighting</b>				
Daylight	1001	10.20%	27.30%	62.50%
Dusk	97	13.40%	27.80%	58.80%
Insufficient lighting	70	21.40%	38.60%	40.00%
Sufficient lighting	338	13.90%	36.70%	49.40%
Without lighting	29	48.30%	24.10%	27.60%
<b>Visibility restricted by</b>				
Buildings	36	5.60%	36.10%	58.30%
Terrain	16	25.00%	37.50%	37.50%
Vegetation	6	33.30%	50.00%	16.70%
Weather conditions	37	21.60%	54.10%	24.30%
Glare	44	22.70%	47.70%	29.60%
Other	73	12.30%	42.50%	45.20%
Without restriction	1323	11.80%	27.50%	60.70%
<b>Time</b>				
Early morning (12–6 am)	68	26.50%	30.90%	42.60%
Morning (6–12 am)	515	13.80%	27.80%	58.40%
Afternoon (12–6 pm)	475	8.20%	25.50%	66.30%
Evening (6–9 am)	394	11.20%	35.80%	53.00%
Night (9–12 am)	83	23.00%	38.50%	38.50%
<b>Shoulder type</b>				
Does not exist	805	10.40%	30.10%	59.50%
<1.5 m	423	17.70%	30.00%	52.30%
[1.5–2.5] m	284	9.50%	28.50%	62.00%
>2.5 m	23	21.70%	34.80%	43.50%
<b>Sidewalk</b>				
Yes	731	14.90%	27.10%	58.00%
No	804	10.20%	32.30%	57.50%
<b>Lane width</b>				
<3.25 m	282	16.30%	44.30%	39.40%
[3.25–3.75 m]	1081	11.80%	24.60%	63.60%
>3.75 m	172	10.50%	38.90%	50.60%

Table 2. Cont.

Variable	No. of Crashes	Fatal Injury	Severe Injury	Minor Injury
<b>Road markings</b>				
Does not exist	59	10.20%	42.40%	47.40%
Separate lanes only	148	11.50%	25.70%	62.80%
Separate lanes and margins	1313	12.60%	29.50%	57.90%
Separate margins	15	13.30%	53.30%	33.40%
<b>Number of injured</b>				
1 injured	1455	12.10%	29.20%	58.70%
2 injured	72	16.60%	43.10%	40.30%
3 injured	7	28.60%	42.80%	28.60%
>3 injured	1	100.00%	0.00%	0.00%
<b>No. of occupants involved</b>				
1 occupant	1408	12.30%	29.30%	58.40%
2 occupants	88	15.90%	36.40%	47.70%
3 occupants	26	3.80%	38.50%	57.70%
>3 occupants	13	23.10%	30.80%	46.10%
<b>Pedestrian infraction</b>				
Not using crossings	181	16.00%	23.80%	60.20%
Crossing unlawfully	245	22.00%	35.10%	42.90%
Other infractions	86	9.30%	18.60%	72.10%
Without infraction	1017	9.70%	30.70%	59.60%
<b>Pedestrian action</b>				
Crossing between vehicles	13	7.70%	23.10%	69.20%
In front of the bus stop	1	0.00%	0.00%	100.00%
Crossing intersection	294	8.50%	30.30%	61.20%
Crossing roadway	586	14.20%	33.40%	52.40%
Crossing road in section	167	13.80%	17.40%	68.90%
Walking on the sidewalk	53	7.50%	20.80%	71.70%
Walking on the road	163	14.70%	38.70%	46.60%
Working on the road	3	0.00%	0.00%	100.00%
Repairing the vehicle	1	0.00%	0.00%	100.00%
Getting on/off the vehicle	2	0.00%	50.00%	50.00%
Road Assistance Service	1	0.00%	0.00%	100.00%
Invading the road running	18	11.10%	27.80%	61.10%
Other	233	12.40%	26.20%	61.40%
<b>Driver infraction</b>				
Distracted driving	280	13.90%	31.40%	54.60%
Not respecting a signal	9	0.00%	33.30%	66.70%
Not respecting a light	25	4.00%	40.00%	56.00%
Not respecting priority	19	0.00%	31.60%	68.40%
Not respecting a crossing	389	8.50%	31.10%	60.40%
Not respecting police	6	16.70%	16.70%	66.60%
Invade opposite direction	2	50.00%	50.00%	0.00%
Incorrectly rotate	1	0.00%	0.00%	100.00%
Reversing wrongly	2	50.00%	50.00%	0.00%
Overtaking unlawful	4	25.00%	50.00%	25.00%
Not keeping distance	1	0.00%	0.00%	100.00%
Prohibited parking	1	100.00%	0.00%	0.00%
Opposite direction	2	0.00%	0.00%	100.00%
Other infraction	242	13.20%	49.20%	37.60%
No infraction	552	14.80%	17.80%	67.40%
<b>Driver speed infraction</b>				
Inadequate speed	107	13.10%	42.10%	44.80%
Exceeding speed	32	65.60%	25.00%	9.40%
Slow circulation	1	0.00%	0.00%	100.00%
No infraction	1395	11.20%	29.00%	59.80%

### 3.2. Injury Severity Analysis Using MNL

One of the common methods applied to model crash severity data is the multinomial logistic regression [32–34]. This method predicts the probability of category membership on a dependent variable based on multiple independent variables. It is an expanded form of binary logistic regression, which permits more than two groups of the dependent variable. The multinomial logistic regression selects one group as the base condition (reference) for the other groups. Then, a contrast of the outcomes of the dependent variable with this reference group is made. The theoretical concept of this statistical

method is described below. The linear function  $Q$  that defines the injury output  $i$  for observation  $n$  is expressed as follows:

$$Q_{in} = \delta_i X_{in} + \eta_{in} \quad (3)$$

where  $\delta_i$  is a vector of computable coefficients,  $X_{in}$  is a vector of discernible features that affect the severity of pedestrian injury sustained by observation  $n$ .  $\eta_{in}$  is an alteration term that takes into consideration unobserved effects. When the alteration terms are distributed independently and are identical to the generalised distribution of extreme values, the multinomial logit models can be expressed as follows [35]:

$$P_n(i) = \frac{\exp(\delta_i X_{in})}{\sum_i \exp(\delta_i X_{in})} \quad (4)$$

## 4. Results and Discussion

### 4.1. Cluster Analysis

As shown in Table 3, pedestrian-vehicle crashes were grouped by variables by using SPSS software. Pedestrian injury severity was considered as a dependent variable with the following three possible classes: slightly injured, severely injured, or fatally injured. Different models of clusters were estimated, from one to ten, for selecting the suitable number of clusters. For further analysis, pedestrian-vehicle crashes data were divided into four clusters. Table 3 shows the clusters profiles. Cluster 1 consists of 19.7% of the sample, cluster 2 consists of 26.6% of the sample, and clusters 3 and 4 consists of 15.5% and 38.2% of the sample, respectively. The characteristics of these four clusters are given below.

**Table 3.** Distribution of variables and membership to each cluster.

Variable	C1	C2	C3	C4
<b>Driver's age</b>				
<18 years old	0.30%	1.20%	1.30%	2.40%
18–30 years old	21.50%	20.30%	18.90%	26.30%
31–64 years old	60.70%	61.30%	63.00%	59.70%
65 and over	17.50%	17.20%	16.80%	11.60%
<b>Pedestrian's age</b>				
<18 years old	21.50%	14.70%	17.20%	21.80%
18–30 years old	14.20%	7.80%	8.40%	10.80%
31–64 years old	32.00%	30.90%	34.90%	30.90%
65 and over	32.30%	46.60%	39.50%	36.50%
<b>Driver's gender</b>				
Male	78.90%	77.20%	79.00%	78.00%
Female	21.10%	22.80%	21.00%	22.00%
<b>Pedestrian's gender</b>				
Male	49.50%	51.00%	41.60%	52.70%
Female	50.50%	49.00%	58.40%	47.30%
<b>Atmospheric factors</b>				
Good weather	87.10%	82.10%	81.00%	87.50%
Light rain	4.60%	11.00%	11.30%	7.80%
Heavy rain	1.30%	4.20%	3.80%	1.90%
Fog	0.40%	0.70%	0.40%	0.00%
Snow	0.00%	0.00%	0.00%	0.60%
Hail	0.00%	0.20%	0.00%	0.00%
Heavy wind	2.00%	0.00%	0.00%	1.00%
Other	4.60%	1.80%	3.50%	1.20%
<b>Day of the week</b>				
Beginning of week (Mon)	14.90%	14.50%	16.80%	14.70%
Weekday (Tue, Wed, Thu)	48.80%	43.40%	45.80%	47.40%
End of week (Fri)	19.50%	18.10%	17.20%	16.20%
Weekend (Sat, Sun)	16.80%	24.00%	20.20%	21.70%
<b>Type of day</b>				
Holiday	10.60%	20.80%	13.00%	17.60%
Working day	71.20%	60.80%	60.90%	64.80%
Eve of holiday	9.90%	9.80%	11.80%	8.70%
Day after a holiday	8.30%	8.60%	14.30%	8.90%

Table 3. Cont.

Variable	C1	C2	C3	C4
<b>Lighting</b>				
Daylight	73.90%	59.80%	69.30%	62.80%
Dusk	5.90%	6.40%	7.10%	6.10%
Insufficient lighting	12.20%	20.60%	16.40%	16.70%
Sufficient lighting	6.00%	11.80%	6.30%	11.80%
Without lighting	2.00%	1.40%	0.90%	2.60%
<b>Visibility restricted by</b>				
Buildings	1.30%	2.90%	1.70%	2.70%
Terrain	0.30%	0.70%	1.70%	1.40%
Vegetation	0.00%	0.50%	0.40%	0.50%
Weather conditions	1.00%	4.90%	2.50%	1.40%
Glare	0.60%	4.40%	2.10%	3.20%
Other	4.60%	2.80%	4.60%	6.70%
Without restriction	92.20%	83.80%	87.00%	84.10%
<b>Time</b>				
Early morning (12–6 am)	4.60%	3.40%	4.20%	5.10%
Morning (6–12 am)	36.30%	35.30%	39.10%	28.70%
Afternoon (12–6 pm)	35.30%	26.50%	29.00%	32.60%
Evening (6–9 am)	20.80%	28.40%	23.10%	27.30%
Night (9–12 am)	3.00%	6.40%	4.60%	6.30%
<b>Shoulder type</b>				
Does not exist	68.30%	43.10%	64.30%	49.10%
<1.5 m	20.20%	24.30%	16.80%	27.50%
[1.5–2.5] m	11.20%	29.90%	17.60%	22.00%
>2.5 m	0.30%	2.70%	1.30%	1.40%
<b>Sidewalk</b>				
Yes	28.40%	70.10%	43.70%	48.50%
No	71.60%	29.90%	56.30%	51.50%
<b>Lane width</b>				
<3.25 m	24.40%	33.80%	21.00%	30.50%
[3.25–3.75 m]	65.70%	60.00%	70.60%	58.20%
>3.75 m	9.90%	8.60%	8.40%	11.30%
<b>Road markings</b>				
Does not exist	2.30%	3.90%	2.10%	5.30%
Separate lanes only	6.20%	9.60%	5.00%	10.10%
Separate lanes and margins	90.80%	84.80%	92.40%	83.80%
Separate margins	0.70%	1.70%	0.50%	0.80%
<b>Number of injured</b>				
1 injured	95.00%	96.80%	96.60%	92.50%
2 injured	4.60%	2.80%	2.90%	6.70%
3 injured	0.40%	0.30%	0.50%	0.80%
>3 injured	0.00%	0.10%	0.00%	0.00%
<b>No. of occupants involved</b>				
1 occupant	93.40%	89.50%	89.90%	93.20%
2 occupants	4.60%	6.60%	7.60%	4.90%
3 occupants	1.30%	2.70%	1.70%	1.20%
>3 occupants	0.70%	1.20%	0.80%	0.70%
<b>Pedestrian infraction</b>				
Not using crossings	16.20%	6.60%	2.10%	22.20%
Crossing unlawfully	12.20%	6.10%	2.10%	33.10%
Other infractions	10.90%	1.00%	6.30%	5.80%
Without infraction	60.70%	86.30%	89.50%	38.90%
<b>Pedestrian action</b>				
Crossing between vehicles	0.00%	1.00%	0.00%	1.50%
In front of the bus stop	0.00%	0.20%	0.00%	0.00%
Crossing intersection	0.00%	21.60%	0.00%	22.00%
Crossing roadway	0.00%	45.10%	0.00%	49.50%
Crossing road in section	0.00%	13.00%	0.00%	11.40%
Walking on the sidewalk	0.00%	6.40%	0.00%	4.60%
Walking on the road	0.00%	12.50%	0.00%	10.60%
Working on the road	0.00%	0.20%	0.00%	0.40%
Repairing the vehicle	0.30%	0.00%	0.00%	0.00%
Getting on/off the vehicle	0.70%	0.00%	0.00%	0.00%
Road Assistance Service	0.00%	0.00%	0.40%	0.00%
Invading the road running	5.30%	0.00%	0.80%	0.00%
Other	93.70%	0.00%	98.80%	0.00%

Table 3. Cont.

Variable	C1	C2	C3	C4
<b>Driver infraction</b>				
Distracted driving	0.00%	33.80%	59.70%	0.00%
Not respecting a signal	0.00%	1.70%	0.80%	0.00%
Not respecting a light	0.00%	4.20%	3.40%	0.00%
Not respecting priority	0.00%	3.40%	2.10%	0.00%
Not respecting a crossing	0.00%	55.10%	33.20%	0.00%
Not respecting police	0.00%	1.00%	0.80%	0.00%
Invade opposite direction	0.00%	0.50%	0.00%	0.00%
Incorrectly rotate	0.00%	0.30%	0.00%	0.00%
Reversing wrongly	0.00%	0.00%	0.00%	0.30%
Overtaking unlawful	0.00%	0.00%	0.00%	0.70%
Not keeping distance	0.00%	0.00%	0.00%	0.20%
Prohibited parking	0.00%	0.00%	0.00%	0.20%
Opposite direction	0.00%	0.00%	0.00%	0.30%
Other infraction	30.40%	0.00%	0.00%	28.30%
No infraction	69.60%	0.00%	0.00%	70.00%
<b>Driver speed infraction</b>				
Inadequate speed	1.70%	13.20%	3.80%	4.40%
Exceeding speed	0.60%	2.00%	1.70%	3.10%
Slow circulation	0.00%	0.20%	0.00%	0.00%
No infraction	97.70%	84.60%	94.50%	92.50%

**Cluster 1.** This group includes 60.7% of the accidents wherein the driver is between 31 and 64 years old. A victim aged 31–64 years old or an elderly pedestrian (>65 years old) is involved in most of the accidents in this group. As can be seen in Table 3, the gender of drivers has been mostly male (78.9%). On the other hand, the gender of the injured pedestrian has been divided into the female and the male with the values of 50.5% and 49.5%, respectively. The collisions occurred on working days in 71.2% of the cases. Moreover, it was observed that these accidents occurred under daylight conditions and without any visibility restriction. Furthermore, most of the accidents of this group occurred in crosstown roads with no shoulder (68.3% of the cases) and with no sidewalk (71.6% of the total accidents of this group). In addition, the lane width that characterises this group has been estimated to be between 3.25–3.75 m (65.7%). As can be seen in Table 2, the users of the roads involved in these accidents have not committed any relevant infraction. Therefore, this group can be defined as ‘pedestrian-vehicle collisions on crosstown roads without sidewalk or shoulder under daytime conditions and with no infractions committed’.

**Cluster 2.** This group includes 61.3% of the accidents wherein the driver is between 31 and 64 years old. Most of the injured pedestrians in this group were elderly pedestrians. Similar to cluster 1, the gender of drivers was mostly male (77.2%), unlike the pedestrians involved, who were divided into the female and the male gender. Moreover, the crashes occurred on weekdays in 43.4% of the cases and on working days in 60.8% of the cases. Furthermore, most of these collisions occurred on crosstown roads with no shoulder or a shoulder of less than 1.5 m. However, 70.1% of these collisions occurred on a sidewalk. With regard to infractions, it is relevant to mention that most of the drivers do not respect a pedestrian crossing (55.1%). In addition, 33.8% of the drivers were distracted and 13.2% of them were driving with an inadequate speed. Therefore, this cluster can be defined as ‘pedestrian-vehicle crashes on crosstown roads with a sidewalk during weekdays and caused by not respecting a crossing and distracted driving’.

**Cluster 3.** This group includes 63.0% of the accidents wherein the driver was between 31 and 64 years old. However, most of the victims were pedestrians of more than 65 years old. Similar to clusters 1 and 2, the gender of drivers was mostly male (79.0%), whereas the gender of pedestrians was mostly female (58.4). Moreover, the accidents occurred on weekdays in 45.8% of the cases. Lighting in most of these crashes was adequate and there were no restrictions with regard to visibility. Furthermore, most of these crashes occurred on crosstown roads without shoulder or sidewalk. With regard to infractions, pedestrians did not commit any infractions in the accidents, but 60% of the drivers were driving distractedly. In addition, the other 33% of the drivers do not respect a pedestrian crossing.

Hence, this cluster can be defined as ‘pedestrian-vehicle accidents on crosstown roads without shoulder or sidewalk, with elderly pedestrians involved and caused by not respecting a crossing and distracted driving’.

**Cluster 4.** This group includes 59.7% of the accidents wherein the driver is between 31 and 64 years old. However, it is important to note that this group involved the highest percentage of young drivers (26.3%). Most of the pedestrians are aged 31–64 years old and also over 65 years old with 30.9% and 36.5%, respectively. Moreover, the collisions occurred on working days in 64.8% of the cases. These accidents occurred under daylight conditions and without a visibility restriction. Furthermore, most of the accidents in this group occurred in crosstown roads with no shoulder (49.1% of the cases) or less than 1.5 m (27.5%) and with no sidewalk (51.5% of the total accidents in this group). In addition, the lane width that characterises this group is estimated to be between 3.25–3.75 m (58.2%), but there is a large group that uses a very narrow lane (30.5%, <3.25 m). Fifty-five percent of pedestrians involved in these accidents crossed unlawfully and did not use crossings. Therefore, this group can be defined as ‘pedestrian-vehicle collisions with a relevant percentage of young drivers and elderly pedestrians, on crosstown roads with no shoulder or sidewalk wherein pedestrians have crossed unlawfully’.

It is important to mention that it would have been ideal to have a greater sample size in order to improve the representativeness of the clusters (including those cases of pedestrians’ crosses in which there are no accidents).

#### 4.2. Injury Severity Analysis Using MNL

The primary objective of this study is to explore the different contributing factors that are responsible for increasing the probability of a fatal outcome considering the fact that a pedestrian-vehicle crash has occurred in Spanish crosstown roads. In this analysis, a multinomial logit model was applied for each cluster and for the whole database, where the pedestrian injury severity was considered as a dependent variable with the following three possible classes: slightly injured, severely injured, or fatally injured. In this model, a total of 20 variables were considered, which includes the age of the drivers involved, the gender of the driver, the age of the pedestrians involved, the gender of the pedestrian, atmospheric factors, the day of the week when the accident occurred, the type of day, lighting conditions, the visibility restrictions during the accident, time, lane width, shoulder type, the presence of sidewalk, the state of the road markings, the total number of injuries, the number of the vehicle occupants involved, the infractions committed by the pedestrian, the pedestrian action, the infractions committed by the driver and the possible speed infractions committed by the driver. Using the maximisation of the log-likelihood method, a total of five models were developed, one for the whole data set and one for each cluster (from clusters 1–4). Moreover, a minor injury crash was selected as the base outcome in all models.

The effects of a contributing variable on the conditional probability of a fatal outcome in the case of a fatal or severe crash compared to a minor crash are shown by the estimated coefficients. Table 4 shows the estimation results of the different models. Following Kim et al. (2007) [34] and Sasidharan et al. (2015) [36], a significance level of 10% was used in this analysis. In Table 4, only statistically significant variables at a significance level of 10% have been represented.



The predictors with positive coefficients indicate an increase in the probability of occurrence of fatal or severe injury crashes as compared to minor injury crashes. Moreover, variables that significantly increase the probability of fatal and severe crashes considering the whole data set model are as follows: visibility restricted by weather conditions or glare, infractions committed by pedestrians such as not using crossings, crossing unlawfully, or walking on the road, infractions committed by the driver such as distracted driving, not respecting a light or a crossing and finally, speed infractions by drivers such as an inadequate speed. The results revealed that the variable pedestrian's age is very significant. The odds ratio estimated for a pedestrian's age between 18 and 30 years old was 0.531 ( $e^{-0.633}$ ). This analysis suggests that the probability of a fatal or severance outcome decreases when the pedestrian is aged between 18 and 30 years old compared to the elderly (aged 65 and over), as well as it happens with the other age groups.

With regard to visibility restrictions, the odds ratio estimated for visibility restricted by weather conditions was 2.654 ( $e^{0.976}$ ) and 2.460 ( $e^{0.900}$ ) for visibility restricted by glare. These results suggest that visibility restrictions increase the conditional probability of a fatal or severe outcome in the case of a crash. On the other hand, it has been demonstrated that other factors such as the lane width had also an important effect on crash severity. The odds ratio estimated for a lane width of fewer than 3.25 m was 0.602 ( $e^{-0.507}$ ) and 0.547 ( $e^{-0.604}$ ) for a lane width of between 3.25–3.75 m, compared to a lane width of more than 3.75 m. These results suggest that a narrower lane decreases the conditional probability of a fatal or severe outcome in the case of a crash involving a pedestrian. At first glance, this result could be expected, since according to previous literature, a wide lane may encourage higher speeds and hence may increase the severity. Something similar happens with the shoulder width. The severity is lower for shoulder width of less than 2.5 m, as it can be seen in Table 4.

It has been observed that meaningful relations can be concealed while analyzing traffic accidents in a large set of heterogeneous data. Many studies have demonstrated that segmenting the data into homogeneous groups helps in reducing heterogeneity and provides further information on traffic safety analysis [27,28]. On the other hand, some of the variables that have not been identified as meaningful in the entire database analysis are considered as determinative for a cluster. As shown in Table 4, the significance of the effects of variables is estimated to be very different in the whole data model and the cluster models. For example, the odds ratio of the age of pedestrians between 18 and 30 years old was determined to be 0.531 ( $e^{-0.633}$ ) for the whole data analysis, while it was determined to be 0.312 ( $e^{-1.164}$ ) for cluster 2. The whole data analysis reports that the odds ratio of a pedestrian (in the age group of 18–30 years old) being involved in a crash of fatal or severe injuries is 46.9% lower than the baseline condition (pedestrians older than 65 years old). Nevertheless, the odds ratio estimated for cluster 2 demonstrated that pedestrians between 18 and 30 years old, who traverse crosstown roads without shoulder, with sidewalk and where drivers have committed driving infractions, are 68.8% less likely to suffer fatal or severe injuries as compared to the baseline condition (minor injuries).

Furthermore, some variables are considered significant only for certain clusters, which provide added information, as can be seen in Table 4. For example, the variable early morning is not considered significant in the whole database, while the odds ratio is estimated to be 0.158 ( $e^{-1.848}$ ) for cluster 2. This indicates that the odds ratio of a pedestrian being involved in a crash with fatal or severe injuries when the accident occurs on crosstown roads without shoulder, with sidewalk and where drivers commit infractions is 84.2% lower in the early morning when compared to the night. Similarly, separate margins marked correctly (compared to no markings) are not considered significant in the whole data analysis. This factor is measured to be 65.30 ( $e^{4.179}$ ) for cluster 4, which suggests that pedestrian-vehicle crashes with young drivers and elderly pedestrians are 6000% more likely to suffer fatal or severe injuries when compared to no separate marking. In other words, crosstown roads with no road markings are more likely to cause fatal or serious injuries during pedestrian crashes.

In general, the statistical analysis also shows that pedestrian-vehicle crashes are more likely to cause fatal or severe injuries when pedestrians commit an infraction (such as not using crossings, invading the road or crossing unlawfully, among others). For instance, crossing unlawfully a crosstown

road is 215% more likely to result in serious injury or fatal injury when compared to the baseline condition (no infraction committed). Similarly, exceeding speed (compared to no infraction) is considered significant in the whole data analysis. This factor is estimated to be 3.48 ( $e^{1.246}$ ), which indicates that where drivers exceed speed, pedestrians injured in accidents are 248% more likely to result in serious injury or fatality when compared to the baseline condition (no speed infraction committed by the driver). With regard to infrastructure factors, the odds ratio for a lane width between 3.25 m and 3.75 m was estimated to be 0.547 ( $e^{-0.604}$ ) and it was 0.602 ( $e^{-0.507}$ ) for a lane width of less than 3.25 m. These results indicate that a wider lane increases the conditional probability of a fatal outcome in the case of a crash. In addition, conditions of poor visibility caused by bad weather or glare increase the severity of pedestrian injuries on crosstown road accidents.

## 5. Conclusions

In order to examine the influence of infractions of pedestrians and drivers, crash and infrastructure factors on the severity of the pedestrian injury, this study applied a *k*-means cluster and a multinomial logit model. From 2006 to 2016, a total of 1535 accidents (involving one vehicle and one pedestrian) were examined on Spanish crosstown roads. Moreover, the severity of injured pedestrians was divided considering the consequences of the accident: fatal (death), serious injury, and minor injury. First, crash data were segmented into a homogeneous group using clustering techniques. The statistical analysis shows that factors such as visibility restricted by weather conditions or glare, infractions committed by the pedestrian (such as not using crossings, crossing unlawfully, or walking on the road), infractions committed by the driver (such as distracted driving, not respecting a light or a crossing) and finally speed infractions by drivers (such as inadequate speed) increase the severity of the pedestrian injured. On the other hand, factors such as the age, the shoulder width (1.5–2.5 m), the existence of pavement and lane width of no more than 3.75 m, the early morning traffic, eve of holiday and the existence of road markings are the variables that are associated with less severe injured victims.

In addition, the results also show that combining clustering techniques such as *k*-means cluster and multinomial logit models can successfully provide the underlying patterns pertaining to accident data. However, variables that are not significant have also been identified and found to be very meaningful in the case of a specific cluster. It can, therefore, be concluded that clustering techniques are found to be a useful tool for segmenting crash data.

Furthermore, infractions committed by drivers and pedestrians have been proved to be determinant (significant) factors in this specific type of road. Infractions such as not using road crossings or crossing unlawfully have been constantly interpreted throughout the analysis of accidents. Moreover, the pedestrian infractions occur due to a subject with physical limitation (elderly or disable pedestrians), or due to negligence (road safety education). In order to better design practical policy measurements for our cities, the behavior of elderly pedestrians should be analysed in more detail. Firstly, pedestrian crossings must be consistent, intuitive and well-marked. Moreover, special attention should be paid to older pedestrians because in Spain the share of the elderly population, aged over 65, is projected to increase. Secondly, traffic rules and regulations should be refreshed and enforced for young drivers and repeat offenders. It is evident that new strategies are required to integrate crosstown roads into the urban structure once they become obsolete. Part of the solution can be redundant signalling and the implementation of alternative traffic calming devices. In addition, the main pedestrian crossings traversing the crosstown road should be analysed taking into consideration the main purpose of the pedestrian trip. Our results demonstrate that daylight condition contributes to less severe consequences on the severity of a pedestrian injury. Hence, it is recommended to increase the level of lighting in order to prevent pedestrian severity. Finally, it is also important to mention the infractions committed by drivers. The most frequent infractions committed by drivers have been distracted driving, not respecting a crossing or a traffic light and driving in the opposite direction. It has also been observed that most of the drivers drive at an inadequate speed. Hence, the results show that driving at an inappropriate speed increases the probability of a pedestrian being involved in more severe injuries.

Hence, a set of strategic action plans at the urban level can be designed based on the above results of the statistical analysis. In order to achieve the objective of zero pedestrian-vehicle accidents, the statistical techniques used in this study can help policymakers of transportation departments to determine crucial crash factors and to implement safety countermeasures.

In conclusion, this paper aimed to provide a first in-depth analysis of pedestrian behaviour in the specific road environment, with the objective of designing safer pedestrian routes and fostering more pedestrian mobility as a sustainable mode of transport. In terms of the methodology, further research is needed into the cluster attributes and patterns identified to determine more fully how these factors can be mitigated to reduce the risk of severe injury. Other variables, such as urban services accessibility index or other territorial variables have to be introduced to the model in order to analyse the impact of social severance on crosstown road accidents. Layout variables or exposure variables could be very relevant and have to be considered in further research. Similarly, other statistical methods such as latent class analysis or interactive tools such as Geographic Information System could also provide help to segment accident data or characterize the density of urban environment, respectively. This study can also be transferred to other types of roads or road scenarios. The conclusions from this research could help policymakers to identify critical crash factors and develop safety countermeasures to reduce pedestrian injuries.

**Author Contributions:** Conceptualization, N.C.-S. and B.G.; methodology, N.C.-S. and A.L.G.; software, N.C.-S.; validation, A.L.G.; formal analysis, N.C.-S., A.L.G., and M.A.; investigation, N.C.-S., B.G., and M.A.; resources, N.C.-S. and B.G.; data curation, A.L.G.; writing-original draft Preparation, N.C.-S. and B.G.; writing-review and editing, N.C.-S. and B.G.; visualization, A.L.G. and M.A.; supervision, B.G.; project administration, B.G.; funding acquisition, B.G.

**Funding:** This research was funded by the Spanish Directorate General of Traffic (DGT), through the project SPIP2017-02199, signed in the year 2018 with the Technical University of Madrid.

**Acknowledgments:** This paper shows the results of research sponsored and funded by the Spanish Directorate General of Traffic (DGT) as part of the grants programme for research projects in the field of traffic, mobility and road safety for the year 2017. This project has been referenced with the administrative code SPIP217-02199. The authors are also grateful to the DGT for providing the data necessary for this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. World Health Organization. Road Traffic Injuries—Fact Sheet. 2018. Available online: <http://www.who.int/mediacentre/factsheets/fs358/en/> (accessed on 26 May 2019).
2. Moudon, A.V.; Lin, L.; Jiao, J.; Hurvitz, P.; Reeves, P. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in king country. *Accid. Anal. Prev.* **2011**, *43*, 11–24. [[CrossRef](#)]
3. European Commission. *Traffic Safety Basic Facts on Pedestrians*; European Commission, Directorate General for Transport, 2018; Available online: [https://ec.europa.eu/transport/road\\_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs20xx\\_pedestrians.pdf](https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs20xx_pedestrians.pdf) (accessed on 14 May 2019).
4. Dirección General de Tráfico. Las Principales Cifras de la Siniestralidad vial. España. 2016. Available online: <http://www.dgt.es/es/seguridad-vial/estadisticas-e-indicadores/publicaciones/principales-cifras-siniestralidad/> (accessed on 26 May 2019).
5. Abellán, A. Road traffic crash data: An overview on sources, problems and collection methods. *J. Transp. Technol.* **2017**, *7*, 206–219. [[CrossRef](#)]
6. Lee, C.; Abdel-Aty, M. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accid. Anal. Prev.* **2005**, *37*, 775–786. [[CrossRef](#)]
7. Kim, J.K.; Ulfarsson, G.F.; Shankar, V.N.; Kim, S. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accid. Anal. Prev.* **2008**, *40*, 1695–1702. [[CrossRef](#)]
8. Ulfarsson, G.F.; Kim, S.; Booth, K.M. Analyzing fault in pedestrian-motor vehicle crashes in North Carolina. *Accid. Anal. Prev.* **2010**, *42*, 1805–1813. [[CrossRef](#)]
9. Abdul-Aziz, H.M.; Ukkusuri, S.V.; Hasan, S. Exploring the determinants of pedestrian-vehicle crash severity in New York City. *Accid. Anal. Prev.* **2010**, *50*, 1298–1309. [[CrossRef](#)]

10. Pour-Rouholamin, M.; Zhou, H. Investigating the risk factors associated with pedestrian injury severity in Illinois. *J. Saf. Res.* **2016**, *57*, 9–17. [[CrossRef](#)]
11. Ivan, J.N.; Gårder, P.E.; Zajac, S.S. *Finding Strategies to Improve Pedestrian Safety in Rural Areas*; Connecticut Transportation Institute, University of Connecticut: Storrs, CT, USA, 2000.
12. Islam, S.; Jones, S.L. Pedestrian at-fault crashes on rural and urban roadways in Alabama. *Accid. Anal. Prev.* **2014**, *77*, 267–276. [[CrossRef](#)]
13. Sze, N.N.; Wong, S.C. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accid. Anal. Prev.* **2007**, *39*, 1267–1278. [[CrossRef](#)]
14. Ballesteros, M.F.; Dischinger, P.C.; Langenberg, P. Pedestrian injuries and vehicle type in Maryland, 1995–1999. *Accid. Anal. Prev.* **2004**, *36*, 73–81. [[CrossRef](#)]
15. Casado-Sanz, N.; Guirao, B. Analysis of the impact of population ageing and territorial factors on crosstown roads safety: The Spanish case study. *Transp. Res. Procedia* **2018**, *33*, 283–290. [[CrossRef](#)]
16. Uddin, M.; Huynh, N. Factors influencing injury severity of crashes involving HAZMAT trucks. *Int. J. Transp. Sci. Technol.* **2018**, *7*, 1–9. [[CrossRef](#)]
17. Wu, Q.; Chen, F.; Zhang, G.; Liu, X.C.; Wang, H.; Bogus, S.M. Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. *Accid. Anal. Prev.* **2014**, *72*, 105–115. [[CrossRef](#)]
18. Haleem, K.; Alluri, P.; Gan, A. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accid. Anal. Prev.* **2015**, *81*, 4–23. [[CrossRef](#)]
19. Rezapour, M.; Ksaibatu, K. Application of multinomial and ordinal logistic regression to model injury severity of truck crashes, using violation and crash data. *J. Mod. Transp.* **2018**, *26*, 268–277. [[CrossRef](#)]
20. Chen, Z.; Fan, W. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *Int. J. Transp. Sci. Technol.* **2019**, *8*, 43–52. [[CrossRef](#)]
21. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. [[CrossRef](#)]
22. Depaire, B.; Wets, G.; Vanhoof, K. Traffic accident segmentation by means of latent class clustering. *Accid. Anal. Prev.* **2008**, *40*, 1257–1266. [[CrossRef](#)]
23. Rodríguez, M.J.; Mora, R. *Análisis de Clúster o Análisis de Conglomerados*; Universidad de Alicante: Alicante, Spain, 2001; pp. 145–155.
24. Fraley, C.; Raftery, A. Model-based clustering, discriminant analysis and density estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [[CrossRef](#)]
25. Pardillo-Mayora, J.; Domínguez-Lira, C.; Jurado-Piña, R. Empirical calibration of a roadside hazardousness index for Spanish two-lane rural roads. *Accid. Anal. Prev.* **2010**, *42*, 2018–2023. [[CrossRef](#)]
26. Karlaftis, M.G.; Tarko, A.P. Heterogeneity considerations in accident modelling. *Accid. Anal. Prev.* **1998**, *30*, 425–433. [[CrossRef](#)]
27. Sun, M.; Sun, X.; Shan, D. Pedestrian crash analysis with latent class clustering method. *Accid. Anal. Prev.* **2019**, *124*, 50–57. [[CrossRef](#)]
28. De Oña, J.; López, G.; Mujalli, R.; Calvo, F. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. *Accid. Anal. Prev.* **2013**, *51*, 1–10. [[CrossRef](#)]
29. Montella, A.; Andreassen, D.; Tarko, A.P.; Turner, S.; Mauriello, F.; Imbriani, L.L.; Romero, M.A.; Singh, R. Critical review of international crash databases and proposals for improvement of the Italian national database. *Procedia Soc. Behav. Sci.* **2012**, *53*, 49–61. [[CrossRef](#)]
30. Utriainen, R.; Pöllänen, M.; Liimatainen, H. Road safety comparisons with international data on seriously injured. *Transp. Policy* **2018**, *66*, 138–145. [[CrossRef](#)]
31. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 1967; Volume 1: Statistics. University of California Press: Berkeley, CA, USA, 1967; pp. 281–297. Available online: <https://projecteuclid.org/euclid.bsm/1200512992> (accessed on 29 May 2019).
32. Ulfarsson, G.; Mannering, F. Difference in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accid. Anal. Prev.* **2004**, *36*, 135–147. [[CrossRef](#)]
33. Islam, S.; Mannering, F. Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence. *J. Saf. Res.* **2006**, *37*, 267–276. [[CrossRef](#)]

34. Kim, K.; Yamashita, E. Using a k-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii. *J. Adv. Transp.* **2007**, *41*, 69–89. [[CrossRef](#)]
35. MacFadden, D. Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications*; Manski, C., McFadden, D., Eds.; MIT Press: Cambridge, MA, USA, 1981.
36. Sasidharan, L.; Wu, K.F.; Menendez, M. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. *Accid. Anal. Prev.* **2015**, *85*, 219–228. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).