DTU Library

Integrating police reports with geographic information system resources for uncovering patterns of pedestrian crashes in Denmark

Prato, Carlo Giacomo; Kaplan, Sigal; Patrier, Alexandre; Rasmussen, Thomas Kjær

[Link back to DTU Orbit](#)

Integrating police reports with geographic information system resources for uncovering patterns of pedestrian crashes in Denmark

Carlo G. Prato [a*], Sigal Kaplan [b], Alexandre Patrier [c], Thomas K. Rasmussen [c]

[a] School of Civil Engineering
The University of Queensland
St. Lucia 4072, Brisbane, Australia

[b] Department of Geography
Hebrew University of Jerusalem
Mount Scopus, 91905 Jerusalem, Israel

[c] Department of Management Engineering
Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

[*] corresponding author
School of Civil Engineering, The University of Queensland
St. Lucia 4072, Brisbane, Australia
Telephone: +61.7.33651569, E-mail: c.prato@uq.edu.au

Integrating police reports with geographic information system resources for uncovering patterns of pedestrian crashes in Denmark

Abstract

Promoting walking goes a long way in contributing to the sustainability and health of future cities and regions, and improving pedestrian safety is essential for building more sustainable and healthier communities. As the problem is multifaceted in nature, this study looks at patterns of pedestrian crashes from a perspective that goes beyond the traditional investigation of pedestrian characteristics and behaviour by analysing the contribution of built environment, land use, and traffic conditions. Moreover, this study goes beyond the traditional analysis of traditional police reports by integrating them with rich geographic information system resources. This study analysed a sample of 7469 crashes between a pedestrian and another road user that occurred in Denmark between 2006 and 2015. The crash locations were geocoded and matched to a detailed traffic network, a transport planning model, and several resources detailing building and land use composition. Latent class analysis uncovered patterns of pedestrian crashes for both the fully identified records and the substantial amount of hit-and-run records. Findings from this study reveal a major red thread in the lack of hazard awareness for both pedestrians and road users and suggest solutions from both the behavioural and the infrastructure perspectives. Major needs are (i) educating pedestrians about the risks related to drinking and then walking along major roads in the darkness, (ii) making crossings for pedestrians and approaches for road users easier to understand and to access in order to reduce unnecessary conflicts, and (iii) designing traffic calming solutions around major shopping and leisure locations in dense city centres.

Key words: pedestrian crashes; GIS resources; built environment; land use; traffic exposure.

1. Introduction

Promoting walking goes a long way in contributing to the future of cities and regions by building healthier and more sustainable communities while reducing traffic and pollution (Tight et al., 2011). Although walking has been considered often as secondary to motorised travel, research into pedestrian behaviour has flourished recently, especially looking at its relation with urban form (Saelens and Handy, 2008; Ewing and Cervero, 2011; Kaplan et al., 2016) and its incorporation into regional planning models (Clifton et al., 2016). It is incontrovertible that there is no sustainability without safety, and that promoting walking requires providing a safe environment to pedestrians. As 22% of the 1.24 million yearly road fatalities worldwide are pedestrians (World Health Organisation, 2013), understanding factors contributing to the occurrence of pedestrian crashes as well as uncovering patterns of pedestrian crashes is ever so needed with the aim of allocating resources to the planning of preventive measures.

Existing research on pedestrian crashes has generally focused on factors that affect crash rates and injury severity. A great deal of attention has been dedicated to age and gender of the pedestrians because of the higher involvement of males (Al-Madani and Al-Janahi, 2006; Kim et al., 2008; Mabunda et al., 2008; Prato et al., 2012; Mohamed et al., 2013; Iragavarapu et al., 2015; Tulu et al., 2015) and the higher vulnerability of children and/or elderly (Al-Ghamdi, 2002; Preusser et al., 2002; Eluru et al., 2008; Prato et al., 2012; Hanson et al., 2013; Abdel-Aty et al., 2013; Tulu et al., 2015). Attention has been given also to the type of vehicle involved in the crash (Ballesteros et al., 2004; Kim et al., 2008), the modality of the crash (Al-Ghamdi, 2002; Preusser et al., 2002; Mohamed et al., 2013; Sasidharan et al., 2015), the location in urban environments and the land use destination (Al-Ghamdi, 2002; Montella et al., 2011; Sasidharan et al., 2015; Amoh-Gyimah et al., 2016; Tran et al., 2016; Wang et al., 2016), and the intoxication level of pedestrians (e.g., Fontaine and Gourlet, 1997; Öström and Eriksson, 2001; Mabunda et al., 2008; Hanson et al., 2013; Iragavarapu et al., 2015).

Existing research on pedestrian crashes has focused extensively on crash frequency models, which allow uncovering the factors that are correlated to a variation in the crash rates, or crash injury severity models, which allow unravelling the factors that aggravate or mitigate the severity of the injuries suffered by the involved road users. Although valuable from the perspective of informing about the relevance of the different factors for crash occurrence and severity outcome, existing studies mostly neglected the multiple facets of the problem and only a few of them looked at the concomitant elements that characterise

pedestrian crashes. This "broad picture approach" translated into performing cluster analysis, but existing efforts fell short by focusing on hotspot identification rather than unravelling the reasons for the crashes (Kim et al., 2007; Dai, 2012), limiting the analysis only to fatal crashes (Prato et al., 2012), and restraining cluster analysis to be a step preliminary to injury severity analysis (Mohamed et al., 2013; Sasidharan et al., 2015). The most notable omission in these studies is however the lack of consideration of the built environment where the crashes occur, specifically not only the infrastructure characteristics that are considered in some studies (Hanson et al., 2013; Mohamed et al., 2013; Sasidharan et al., 2015), but also the land use and traffic characteristics at the crash location that are sparingly considered in other studies (Tulu et al., 2015; Amoh-Gyimah et al., 2016; Tran et al., 2016; Wang et al., 2016).

Although this study recognises the previous use of geographical resources for crash data analysis (Dai, 2012; Hanson et al., 2013; Tulu et al., 2015; Amoh-Gyimah et al., 2016; Tran et al., 2016; Wang et al., 2016), its contribution to the literature is threefold. Firstly, this study provides a comprehensive view of crash patterns by exploiting geocoded police records being matched to a large number of geographic information system (GIS) resources that detail to a great extent the environment where pedestrians walk and crashes occur. From a modelling perspective, it is essential not to incur in the omitted variable bias that often characterises crash frequency and injury severity studies. Secondly, this study frames the characteristics of the infrastructure, built environment, land use and traffic characteristics at the crash locations according to the "Six Ds" theory. From a sustainability perspective, it is essential to understand the role that built environment, land use and traffic characteristics play alongside the road user and crash characteristics. Thirdly, this study takes on the challenge of analysing not only pedestrian crashes where all parties involved are identified, but also hit-and-run crashes. From a crash analysis perspective, it is essential to comprehend where these crashes take place and what factors should be mitigated in order to limit this phenomenon.

With the aim of uncovering patterns of pedestrian crashes, this study analysed 7469 pedestrian crashes occurred in Denmark between 2006 and 2015. The crashes were geocoded in police reports collected by the Danish Road Directorate and were matched to GIS resources containing rich information about built environment, land use and traffic characteristics at the crash locations. Pattern recognition was performed via Latent Class Analysis (LCA) that offered an underlying statistical model, the ability to represent similarity across clusters, and the availability of goodness-of-fit criteria (Lanza et al., 2007).

The remainder of this paper is structured as follows. Initially, section 2 describes data and methods to uncover crash patterns. Then, section 3 presents the results from the LCA. Lastly, section 4 discusses the major findings and touches upon recommendations for preventive measures.

2. Data

2.1 Crash data

Police reports of crashes were retrieved from the national database maintained by the Danish Road Directorate that consists of three separate files: (i) the crash file provides details about location geographical coordinates, collision type, road user maneuvers, involved vehicles and road users, infrastructure characteristics, and environmental conditions; (ii) the person file provides details about demographics, alcohol or drug consumption, restraint use, license validity and injury severity; (iii) the vehicle file provides details about make and model, registration date, maneuver prior to the crash, collision point, and reported damage.

This study analysed crashes involving a pedestrian and a vehicle that occurred on Danish roads during the ten-year period between 2006 and 2015. Crashes involving a pedestrian and a vehicle reduce the noise that considering multiple vehicles and their related dynamics would add (Mohamed et al., 2013; Sasidharan et al., 2015). Matching the crashes to the GIS resources in the version of the year of each crash limited the variation within the ten-year period and guaranteed a large sample size. The crash database contained 7469 records whose characteristics are presented in table 1, and whose outcome ranged over (i) damage only, (ii) light injuries that required proper medical treatment, (iii) severe injuries that involved temporary or permanent incapacitation, and (iv) fatalities that occurred within 30 days of the crash.

A couple of considerations apply to the crash database. Firstly, one group of crashes differed from the others in that the other road user fled the crash scene before the police arrived: as the variables were different because of the missing information on the other road user involved and the literature was non-existent about hit-and-run pedestrian crashes, the analysis focused on two databases of 6539 fully identified and 930 hit-and-run records. Secondly, the geocoded crash locations were matched to the GIS resources to retrieve and verify the characteristics of infrastructure, built environment, land use and traffic conditions according to the GIS databases described in the following section.

[Insert table 1 about here]

2.2 GIS data

GIS resources were used for quantifying the infrastructure, built environment, land use and traffic characteristics at the crash locations according to the "Six Ds" theory (Ewing and Cervero, 2010). Design, density, destination accessibility, distance to transit, diversity, and demand were measured by integrating the crash data with the Danish register of buildings, the Danish register of businesses, the Danish population in 1ha grid cells, the land cover dataset of the European Environment Agency, and the network and traffic of the Danish National Transport Model (NTM) for private and public transport.

Design was measured in terms of road characteristics: category, directions, lanes, section type, intersection type, and speed limits. Density was quantified at the street buffer level (within 500m of the crash location) and the transport zone level (within the traffic zone from the NTM) in order to grasp the effect of both the immediate walking surroundings and the overall traffic level. At the street buffer level, density was measured as land use destination, road density, road being low speed, road being low volume, and intersection number. At the transport zone level, density was measured as population composition, income distribution, car ownership, job density, and unemployment rate. Destination accessibility was assessed as the presence of schools and leisure destinations in the immediate walking surroundings. Distance to transit was expressed as the number of public transport stops or stations in proximity to the crash location. Diversity was measured at the municipality level in terms of region, rurality index (Danish Ministry of Food Agriculture and Fisheries 2011), and land use diversity according to the Herfindahl-Hirschman (HH) index (Song et al., 2013). Demand was retrieved from the NTM in terms of road and walking traffic density in the zone where the crash occurred.

Given the geocoded crash locations, the following measures (presented in table 2) were calculated to follow the "Six Ds" theory: (i) measures of design were extracted from the NTM road network in order to correct for incomplete records in the crash data, with the network being updated over the years and hence the information being current with the crash year; (ii) measures of density, destination accessibility and distance to public transport for the street buffer level were evaluated within 500m of the crash locations at the aerial level from the register of buildings, the land cover dataset and the NTM network (500m is the 90th percentile of walking trips in Denmark); (iii) measures of density, diversity and demand at the transport zone level were determined from the NTM network. The NTM zones represent homogeneous portions of neighborhoods around the crash locations: the traffic zone layer and

the high-definition road network allowed retrieving details about the road characteristics as well as the traffic for both vehicles and pedestrians.

[Insert table 2 about here]

3. Method

LCA was applied to the joint crash-GIS database and was selected over competing clustering techniques for theoretical and practical reasons. Theoretically, LCA is to be preferred because of its reliance on an underlying statistical model, its ability to predict cluster probabilities for new records, and its provision of goodness-of-fit criteria for the decision about the number of clusters (Lanza et al., 2007; Depaire et al., 2008). Practically, LCA is to be preferred because of its demonstrated suitability for the analysis of crash data with particular emphasis on cyclists (Kaplan and Prato, 2013) and pedestrians (Mohamed et al., 2013; Sasidharan et al., 2015).

LCA classifies similar records into K latent classes and, given a vector of N crashes defined by M variables ($y_i = y_1, \ldots, y_M$) and a vector Y_i ($Y_i = Y_{i1}, \ldots, Y_{iM}$) of values of the M variables for each crash i, the LCA model is formulated as (Lanza et al., 2007):

$$p\left(Y_i|\theta\right) = \sum_{k=1}^{K} P\left(C_k\right) p\left(Y_i|C_k, \theta_k\right) \tag{1}$$

where k indicates a latent class, $P(C_k)$ is the prevalence of latent class C_k, $p(Y_i | C_k, \theta_k)$ is the conditional multivariate probability that a crash in class C_k would be characterized by Y_i, and θ_k is a vector of parameters to be estimated.

Simplifying assumptions make for an LCA estimable model formulation with reasonable parametric complexity: (i) every variable is transformed in an ordinal variable with R_m possible responses ($r_{mi} = 1, \ldots, R_{mi}$) that is in agreement with the nature of the variables in the crash database; (ii) the records are assumed to be uncorrelated, which is reasonable for cross-sectional datasets of seemingly unrelated events such as road crashes; (iii) the categorical indicators are assumed to be independent within a latent class, which maintains a parsimonious model structure; (iv) the categorical variables are assumed to be endogenous indicators of the latent class, which implies that no covariates are used to predict the class membership. Under these reasonable assumptions, the LCA model is written as (Lanza et al., 2007):

$$p\left(Y_i|\theta\right) = \sum_{k=1}^{K} \pi_k \prod_{m=1}^{M} \prod_{r_m=1}^{R_m} \theta_{mr_m|k}^{I\left(y_{im}=r_m\right)} \tag{2}$$

where *I* is an indicator function that equals 1 if y_{im} equals r_m and 0 otherwise, and the class membership probabilities π_k and the indicator response probabilities θ_{mr_m} are parameters to be estimated.

The LCA parameters were estimated by maximum likelihood using the two-step expectation-maximisation (EM) algorithm (Lanza et al., 2007): (i) the E-step calculates the expected value of the log-likelihood function with respect to the conditional distribution of *k* given *Y* under the current estimate of the vector θ_k of parameters; (ii) the M-step finds the parameters that maximise the conditional distribution of *k*. Convergence was defined as reaching a maximum absolute deviation less than 1E-06 and, given the risk for finding local maxima, the LCA was performed with 50 sets of random starting values to increase the likelihood of convergence to the global maximum. Estimation was performed by using a SAS procedure (Lanza et al., 2007) and the number of clusters was determined according to the Bayesian Information Criterion (BIC) because of its superiority in terms of consistency and accuracy (Lanza et al. (2007).

4. Results

LCA elicited from the data the crash patterns while using the entire spectrum of the road user and crash characteristics as well as the "six D's" measures. Initially, the analysis separated fully identified from hit-and-run records as the latter did not include information about the other road users and the large amount of missing data could have created problems in the modelling. Then, the analysis included all variables with the exception of pedestrian injury severity (an outcome and not a characteristic of the crash) and tested a variable number of clusters in order to retrieve the optimal number from goodness-of-fit measures. Last, the identification of the clusters looked for over- or under-representation of a characteristic with respect to its distribution in the sample via a Chi-square test with p=0.05 (Weiss et al., 2016).

4.1 Fully identified records

LCA estimation for a number of clusters varying from 2 to 20 revealed a decreasing trend in the BIC values until the solution with 8 clusters where the BIC was approximately constant. As the 8 cluster solution had the highest entropy value (0.95) indicating excellent class separation and homogeneity of the clusters, it was selected as the best solution.

When looking at the distribution of the variables in the fully identified records, several characteristics were prevalent across the 8 clusters: most crashes occurred on weekdays in good weather conditions with dry road surface, on two-way undivided roads with one lane per direction. Interestingly, the other road user was male in about three every

four crashes, was usually driving a car, and intoxication was recorded in about 3 out of 100 road users but in slightly more than 1 every 10 pedestrians.

The interpretation of the 8 clusters relied on observing the distribution of the characteristics for the variables that had a significant difference with respect to their distribution in the sample for at least one of the clusters (details about the LCA cluster compositions are presented in Appendix A1). Figure 1 illustrates the characteristics of the 8 clusters with a visual approach meant to highlight their similarities and differences.

[Insert figure 1 about here]

The location of the crash in urban or rural areas is significantly different from the sample for all clusters, and specifically the first seven clusters have the prevalence of occurring in an urban environment. Clusters C1 through C5 share the prevalent characteristics of occurring mainly in road sections with the other road user going straight and not having the duty to give way, while clusters C6 and C7 share the prevalent characteristics of occurring mainly in intersections where road users are turning, pedestrians are crossing, the location is major local roads where it is dark but illuminated, and there is a high public transport accessibility.

Clusters C1 through C3 share an over-representation of pedestrians crossing minor local roads. Cluster C1 (19.9%) has an over-representation of crashes where the pedestrian crosses from a hidden place, the area is mainly high-rise residential, and the traffic density of both pedestrians and vehicles is medium. Cluster C2 (14.6%) has an over-representation of crashes where elderly pedestrians cross road sections in municipalities that are classified as intermediate or rural and are characterised by low-rise residential areas, low population density, low traffic levels for vehicles, and medium traffic levels for pedestrians. Cluster C3 (9.6%) has an over-representation not only of elderly but also of younger pedestrians, as well as crashes where pedestrians cross road sections in rural or peripheral municipalities mainly in Jutland, in shopping and residential areas with low population density and low levels of vehicle and pedestrian traffic.

Clusters C4 and C5 have an over-representation of intoxicated road users while being differentiated by the fact that pedestrians are mainly crossing in cluster C4, and are mainly standing on or walking along the road in cluster C5. Cluster C4 (5.4%) has an over-representation of crashes where pedestrians cross from hidden locations on two-way divided roads, the other road users ride a bicycle, are intoxicated, and are between 20 and 39 years old, and the location is mainly Copenhagen with high public transport accessibility, a high number of leisure locations, high population density, and high levels of vehicle and

pedestrian traffic in shopping and residential areas. Cluster C5 (6.7%) has an over-representation of crashes where road users are males and under 19 years old, stand or walk on medium to minor local roads in residential areas with low to medium population density and medium levels of vehicle and pedestrian traffic.

Clusters C6 and C7 have an over-representation of crashes where pedestrians cross intersections in major local roads. Cluster C6 (17.9%) has an over-representation of crashes involving pedestrians between 20 and 39 years old and male road users driving mainly a car in Copenhagen, and occurring mainly in shopping and residential areas with high traffic for both vehicles and pedestrians. Cluster C7 (15.5%) has an over-representation of crashes involving female road users turning in green light mainly at 4-legged intersections, and occurring in bad weather and wet road surface that result in poor visibility on two-way divided roads with medium levels of pedestrian and vehicle traffic in areas that are quite dishomogenous from a land use perspective.

The last cluster C8 (10.4%) is very different in its characteristics as it has an over-representation of crashes occurring in road sections in rural areas, with drivers going straight and pedestrians standing on or walking along national or regional roads. These roads have a high speed limit and there is an over-representation of male intoxicated pedestrians, crashes in the weekend and at night, and absence of illumination in the darkness that results in poor visibility. Most crashes occur in rural or peripheral municipalities in Jutland where there is low public transport accessibility and low vehicle and pedestrian traffic.

The pedestrian injury severity was compared across clusters post-estimation. Clearly, cluster C8 was associated with the most severe outcomes as 18.2% of the crashes were fatal and 38.6% of the crashes had severe pedestrian injuries. Also cluster C3 had a higher than average share of serious injuries and fatalities, while clusters C4 and C5 had the less severe consequences as fatalities and serious injuries were under-represented. The location of the clusters was observed post-estimation as illustrated in figure 2. Clearly, crashes in clusters C1, C4 and C6 were located mainy in the four largest cities in Denmark (i.e., Copenhagen, Aarhus, Aalborg, Odense), with the first occurring mainly in the suburbs and the latter two in the inner parts of the cities. Crashes in cluster C2 were mostly located in smaller cities like Køge, Viborg or Fredericia, while crashes in cluster C3 were evenly spread over the country and far from city centres. Crashes in cluster C7 were mainly located in the suburbs of major and medium-sized cities in Denmark (e.g., Copenhagen, Aarhus, Aalborg, Odense, Esbjerg, Fredericia, Viborg, Svendborg). Lastly, crashes in cluster C8 were spread across the nation although a prevalence in Jutland was observed.

4.2 Hit-and-run records

LCA estimation with a number of clusters varying from 2 to 20 revealed a clear minimum for the 5 cluster solution that had also entropy equal to 0.95 suggesting excellent class separation and homogeneity of the clusters.

When observing the distribution of the variables in the hit-and-run records, several characteristics were prevalent across the clusters: most crashes occurred on weekdays in daylight, with good weather conditions, with dry road surface, on two-way undivided roads with one lane per direction. Witnesses mentioned that most of the vehicles that did not stop after the crashes were cars.

The interpretation of the 5 clusters relied on comparing the distribution of the characteristics for the variables with respect to their distribution in the sample (details about the LCA cluster compositions are presented in Appendix A2). Figure 3 illustrates the characteristics of the 5 clusters with a visual approach that emphasises their similarities and differences.

[Insert figure 3 about here]

As for the fully identified records, the location of the crash in urban or rural areas is significantly different for all clusters, and specifically the first four clusters include crashes occurring prevalently in an urban environment. Clusters C1 and C2 share the prevalent characteristics of occurring in road sections with road users going straight, while clusters C3 and C4 share the over-representation of crashes where pedestrians are crossing intersections in urban areas with medium levels of population density, pedestrian traffic and vehicle traffic.

Cluster C1 (25.3%) has an over-representation of crashes where a cyclist was involved on major local roads in the evening or at night and the location was Copenhagen, with several leisure locations, high public transport accessibility, in shopping and residential areas with high population density and high levels of traffic for both pedestrians and vehicles. Cluster C2 (22.1%) has an over-representation of crashes involving pedestrians under 19 years old, who are standing on or walking along medium to minor local roads that are not divided by a median and have one lane per direction, and occurring in low-rise residential areas with low population density and low levels of vehicle and pedestrian traffic.

Cluster C3 (26.8%) has an over-representation of pedestrians crossing from hidden places in national or regional roads where the other road user is going straight, the public transport accessibility is very good, and the location is mainly in Copenhagen or Zealand. Cluster C4 (18.4%) has an over-representation of crashes where pedestrians are female, road

users are turning at intersections between two-way undivided roads, and the location is an intermediate to rural municipality mainly in Zealand and North- or Mid-Jutland.

Lastly, cluster C5 (7.4%) is different from the previous four clusters in this sample, but similar to cluster C8 of the fully identified records. Namely, cluster C5 has an over-representation of crashes occurring in road sections in rural areas, with road users on cars or motorcycles going straight and pedestrians standing on or walking along national or regional roads. These roads have a high speed limit and there is an over-representation of male pedestrians, crashes in the weekend and in the evening or at night, and absence of illumination in the darkness. Most crashes occur in rural or peripheral municipalities in Jutland where there is low vehicle and pedestrian traffic.

The distribution of the pedestrian injury severity in the clusters was observed post-estimation. Unlike for the fully identified records, there was not a significant difference across clusters in terms of pedestrian injury severity as emerged by a chi-square test comparing the distributions. In particular, there were no fatalities in cluster C5, a striking difference from cluster C8 in the previous sample. In general, hit-and-run clusters are not different in terms of severity outcome. The crash location in the clusters was observed post-estimation as depicted in figure 4. Crashes in clusters C1 and C3 were located mainy in the four largest cities in Denmark (i.e., Copenhagen, Aarhus, Aalborg, Odense), with the former in the inner parts of the cities and the latter in the suburbs. Crashes in cluster C2 were spread over the country in less populated areas, while crashes in cluster C4 were mainly located in smaller cities in Denmark (e.g., Køge, Herning or Randers. Lastly, crashes in cluster C5 were spread across the nation with a slight prevalence in North- and Mid-Jutland.

[Insert figure 4 about here]

5. Discussion and conclusions

This study integrated police reports with GIS resources and elicited patterns of pedestrian crashes that allow looking at the problem of collisions between pedestrians and other road users from a broad perspective, rather that the narrow one often taken in crash frequency and injury severity models.

This study confirms previous findings about the relevance of the involvement of young and elderly pedestrians (e.g., Al-Ghamdi, 2002; Preusser et al., 2002; Prato et al., 2012; Hanson et al., 2013), the level of intoxication in the pedestrians (e.g., Öström and Eriksson, 2001; Mabunda et al., 2008; Hanson et al., 2013), and the location in urban areas (e.g., Al-Ghamdi, 2002; Montella et al., 2011). Unlike previous findings, this study shows

equal involvement of males and females rather than higher involvement of males (Al-Madani and Al-Janahi, 2006; Kim et al., 2008; Mabunda et al., 2008; Prato et al., 2012). Moreover, this study adds to the literature perspectives such as the large involvement of male road users and in general the recurrence of certain crashes in locations with a specific land use in the immediate surroundings and a specific traffic pattern in the larger area.

When looking at the consequences, a major problem emerges for rural areas. Looking at the findings from clusters C8 in the fully identified records and C5 in the hit-and-run records, behavioural and infrastructure solutions may be opportune. It seems reasonable that campaigns could address problematic behaviour such as walking intoxicated at night along major roads and more broadly advise that drinking does not affect only driving, but also walking. It seems also reasonable that interventions could increase illumination in darkness, separation between pedestrians and other road users alongside major roads, and in general visibility for hazard recognition and conflict mitigation.

When looking at the prevalence of crash patterns, a second major problem emerges in urban areas. For rural or peripheral municipalities with low traffic of vehicles and pedestrians where in particular young and elderly pedestrians appear more affected, it seems plausible that campaigns could address the lack of awareness about the hazard inherent in not being able to anticipate what could happen (for the young) or to react properly (for the elderly). For dense municipalities with high traffic of vehicles and pedestrians, it seems reasonable to look at solutions that will make safe crossings more accessible, as two clusters have over-representation of crossing from hidden locations in urban areas, will propose traffic calming solutions around major shopping and leisure locations, as two clusters have over-representation of crashes in densely populated and trafficated urban areas, and look at specific issues with bicycles and public transport accessibility.

A third major problem concerns specifically the high percentage of male road users involved in crashes with pedestrians. It is puzzling that three out of four road users are male, and it is clear that there is at least one pattern in which they are male, very young, intoxicated and collide with pedestrians on medium to minor roads where there is not much pedestrian traffic. As the previous two problems, this third one leads to a general lack of awareness about hazards in areas where there is not a lot of pedestrian traffic. When considering the phenomenon of safety in numbers, the increase in awareness where the numbers are not there seems paramount.

A limitation to the study should be recognised in the complexity of merging a large number of data sources that rely on the accurate geo-coding of the crash and the different

definitions of buildings and roads. However, this is also one of the merits of this study that has gathered a significant amount of resources that are becoming more and more accessible and has exploited them for a study that has uncovered the lack of awareness, albeit on different layers, as the major issue that behavioural and infrastructure solutions are required to tackle. Future research could look into exploiting the same resources to uncover factors that aggravate or mitigate injury severity and crash rates, possibly while considering spatial correlation across crashes also to individuate hotspots that require intervention.

References

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. *Transportation Research Part A: Policy and Practice*, 49, 62-75.

Al-Ghamdi, A.S., 2002. Pedestrian-vehicle crashes and analytical techniques for stratified contingency tables. *Accident Analysis and Prevention*, 34, 205-214.

Al-Madani, H., Al-Janahi, A., 2006. Personal exposure risk factors in pedestrian accidents in Bahrain. *Safety Science*, 44, 335-347.

Amoh-Gyimah, R., Sarvi, M., Saberi, M., 2016. Investigating the effects of traffic, socioeconomic, and land use characteristics on pedestrian and bicycle crashes: A case study of Melbourne, Australia. *Proceedings of the 95th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ballesteros, M.F., Dischinger, P.C., Langenberg, P., 2004. Pedestrian injuries and vehicle type in Maryland 1995-1999. *Accident Analysis and Prevention*, 36, 73-81.

Clifton, K.J., Singleton, P.A., Muhs, C.D., Schneider, R.J., 2016. Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, 52, 111-122.

Dai, D., 2012. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. *Journal of Transport Geography*, 24, 206-214.

Danish Government, 2013. The Danish Rural Development Programme 2007-2013. Danish Ministry of Food, Agriculture and Fisheries, Copenhagen, Denmark, available at: http://agrifish.dk/rural_development.aspx?ID=46534, 2011.

Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. *Accident Analysis and Prevention*, 40, 1257-1266.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, 40, 1033-1054.

Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76, 265-294.

Hanson, C.S., Noland, R.B., Brown, C., 2013. The severity of pedestrian crashes: an analysis using Google Street View imagery. *Journal of Transport Geography*, 33, 42-53.

Iragavarapu, V., Khazraee, S.H., Lord, D., Fitzpatrick, K., 2015. Pedestrian fatal crashes on freeways in Texas. *Proceedings of the 94th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Kaplan, S., Nielsen, T.A.S., Prato, C.G., 2016. Walking, cycling and the urban form: A Heckman selection model of active travel mode and distance by young adolescents. *Transportation Research Part D: Transport and Environment*, 44, 55-65.

Kaplan, S., Prato, C.G., 2013. Cyclist-motorist crash patterns in Denmark: A latent class clustering approach. *Traffic Injury Prevention*, 14, 725-733.

Kim, K., Brunner, I.M., Yamashita, E.Y., 2008. Modeling fault among accident-involved pedestrians and motorists in Hawaii. *Accident Analysis and Prevention*, 40, 2043-2049.

Kim, K., Yamashita, E.Y., 2007. Using a K-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii. *Journal of Advanced Transportation*, 41, 69-89.

Lanza, S.T., Collins, L.M., Lemmon, D.R., Schafer, J.L., 2007. PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling*, 14, 671-694.

Mabunda, M.M., Swart, L.A., Seedat, M., 2008. Magnitude and categories of pedestrian fatalities in South Africa. *Accident Analysis and Prevention*, 40, 586-593.

Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, 54, 27-37.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2011. Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record*, 2237, 107-116.

Öström, M., Eriksson, A., 2001. Pedestrian fatalities and alcohol. *Accident Analysis and Prevention*, 33, 173-180.

Prato, C.G., Gitelman, V., Bekhor, S., 2012. Mapping patterns of pedestrian fatal accidents in Israel. *Accident Analysis and Prevention*, 44, 56-62.

Preusser, D.F., Wells, J.K., Williams, A.F., Weinstein, H.B., 2002. Pedestrian crashes in Washington, DC and Baltimore. *Accident Analysis and Prevention*, 34, 703-710.

Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. *Medicine and Science in Sports Exercise*, 40, S550-S566.

Sasidharan, L., Wu, K-F., Menendez, M., 2015. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. *Accident Analysis and Prevention*, 85, 219-228.

Song, Y., Merlin, L., Rodriguez, D., 2013. Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 42, 1-13.

Tight, M., Timms, P., Banister, D., Bowmaker, J., Copas, J., Day, A., Drinkwater, D., Givoni, M., Gühnemann, A., Lawler, M., Macmillen, J., Miles, A., Moore, N., Newton, R., Ngoduy, D., Ormerod, M., O'Sullivan, M., Watling, D., 2011. Visions for a walking and cycling focussed urban transport system. *Journal of Transport Geography*, 19, 1580-1589.

Tran, T., Truong, T., Meyer, T., 2016. Characterizing pedestrian crashes along an urbanization gradient: A case study in Greensboro, North Carolina. *Proceedings of the 95th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Tulu, G.S., Washington, S., Haque, M.M., King, M.J., 2015. Investigation of pedestrian crashes on two-way two-lane rural roads in Ethiopia. *Accident Analysis and Prevention,* 78, 118-126.

Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016. Macro-level safety analysis of pedestrian crashes in Shanghai, China. *Accident Analysis and Prevention*, 96, 12-21.

Weiss, H.B., Kaplan, S., Prato, C.G., 2016. Fatal and serious road crashes involving young drivers in New Zealand: a latent class clustering approach. *International Journal of Injury Control and Safety Promotion*, 23, 427-443.

World Health Organisation, 2013. *The World Health Report 2013: Research for Universal Health Coverage*. WHO Press, Geneva, Switzerland.

Table 1. Crash characteristics

Variable	Category	%	Category	%
	Characteristics of the pedestrian			
Injury severity	No injury	46.1	Severe injury	31.3
	Light injury	17.8	Fatality	4.8
Gender	Male	46.9	Female	53.1
Age	<= 9 years old	6.9	40-49 years old	11.4
	10-14 years old	7.7	50-59 years old	10.4
	15-19 years old	11.4	60-69 years old	9.2
	20-29 years old	16.0	>= 70 years old	16.4
	30-39 years old	10.6		
Intoxicated	No	89.4	Yes	10.6
Behaviour	Crossing from the right	36.5	Walking on the pavement	5.7
	Crossing from the left	26.4	Standing on the road	11.0
	Hidden	8.6	Hit by reversing vehicle	5.9
	Walking along the road	5.9		
Nationality	Danish	87.4	Foreigner	12.6
	Characteristics of the other road user			
Vehicle	Bicycle	6.6	Van	4.7
	Moped/motorcycle	7.9	Heavy vehicle	2.3
	Car	74.6	Bus	3.9
Gender [a]	Male	73.1	Female	26.9
Age [a]	<= 17 years old	2.9	40-49 years old	19.1
	18-19 years old	5.5	50-59 years old	16.3
	20-24 years old	11.0	60-69 years old	10.5
	25-29 years old	8.9	>= 70 years old	7.3
	30-39 years old	18.5		
Intoxicated [a]	No	96.8	Yes	3.2
Driving license [a]	No license required	9.9	Invalid license	3.9
	Valid license	86.2		
Licensing years [a]	0-1 years	7.0	10-19 years	19.3
	2-4 years	7.2	More than 20 years	43.8
	5-9 years	9.4	Not applicable	13.3
Behaviour [a]	Straight	70.5	Reversing	5.9
	Turn right	5.5	Jump off the road	5.6
	Turn left	12.5		
Traffic control [a]	Traffic light: green	15.2	Stop or give way sign	13.5
	Traffic light: red	2.5	No duty to give way	68.8
Nationality [a]	Danish	90.7	Foreigner	9.3
	Characteristics of the crash			
Season	Winter	25.2	Summer	20.6
	Spring	22.3	Autumn	31.9
Type of day	Weekday	78.6	Weekend	21.4
Time of day	9pm – 6am	16.0	3pm – 6pm	26.5
	6am – 9am	13.1	6pm – 9pm	12.8
	9am – 3pm	31.6		
Light	Daylight	61.2	Darkness	6.4
	Twilight	5.2	Illumination	27.2
Visibility	Good visibility	92.8	Poor visibility	7.2
Weather	Good	81.6	Bad	18.4
Road surface	Dry	67.5	Wet/slippery	32.5

Note: [a] characteristics of the 6539 complete records, excluding the 930 hit-and-run records

Table 2. GIS data characteristics

Variable	Category	%	Category	%
	Design measures			
Road category	Expressway and motorway	1.0	Major local	37.2
	National	12.6	Medium local	16.0
	Regional	12.9	Minor local	20.3
Road directions	One-way	6.1	Two-way divided	12.7
	Two-way undivided	81.2		
Lanes per direction	One	92.3	Two or more	7.7
Location	4-legged intersection	20.9	Straight road	57.5
	3-legged intersections	10.7	Curve	1.7
	Roundabout	1.3	Bicycle lane or path	2.7
	Other intersection type	5.1		
Speed limit	<= 50 km/h	6.0	50-80 km/h	10.6
	50 km/h	74.8	80-130 km/h	8.7
	Density measures at the street buffer level (within 500m of the crash location)			
Roadside land use	Shopping street	17.6	Low residential area	19.3
	Industrial area	6.3	Buildings with no access	6.4
	High residential area	35.8	Spare buildings	14.6
Road density	< 10 km/km²	22.0	> 15 km/km²	39.1
	10-15 km/km²	38.9		
Intersection number	<= 25	18.0	51-75	34.8
	26-50	32.4	> 75	14.8
Low speed road (< 30 km/h)	0.0 %	25.0	10.1-25.0 %	26.5
	0.1-10.0 %	36.6	> 25.0 %	11.9
Low volume road (< 2000 veh/day)	No	93.6	Yes	6.4
	Density measures at the zone level (per NTM zone)			
Population density	< 100 inhabitants/km²	14.9	1,000-5,000 inhabitants/km²	40.1
	100-1,000 inhabitants/km²	25.2	> 5,000 inhabitants/km²	19.8
Share of young population	< 20 %	31.3	> 30 %	6.4
	20-30 %	62.4		
Share of elderly population	< 10 %	13.0	> 20 %	17.2
	10-20 %	69.8		
Households without car	< 30 %	15.8	> 50 %	28.0
	30-50 %	56.2		
Job density	<100 jobs/km²	22.6	1,000-5,000 jobs/km²	28.1
	100-1,000 jobs/km²	37.8	> 5,000 jobs/km²	11.5
Average income	< 180,000 DKK/year	21.5	220,000-300,000 DKK/year	15.8
	180,000-220,000 DKK/year	59.8	> 300,000 DKK/year	2.9
Unemployment rate	< 4%	12.2	8-10 %	19.8
	4-8%	52.2	> 10%	15.8
	Destination accessibility measures			
Schools	No	59.2	Yes	40.8
Leisure locations	0	23.4	> 10	36.5
	1-10	40.1		
	Distance to transit measures			
Public transport stops / stations	0	8.9	> 5	31.4
	1-5	59.7		
	Diversity measures			
Region	Copenhagen	37.5	North Jutland	7.6
	Zealand	12.2	Mid-Jutland	20.4
	Fyn	9.4	South Jutland	12.9
Area	Urban	88.1	Rural	11.9
Rural index of the municipality	Urban	56.6	Rural	22.1
	Intermediate	15.0	Peripheral	6.3
HH index	< 0.50	39.7	0.75-0.99	17.9
	0.50-0.75	25.6	1.00	16.8
	Demand measures (per NTM zone)			

Road traffic Density	< 5,000 veh-km/day/km²	23.7	25,000-50,000 veh-km /day/km²	22.1
	5,000-10,000 veh-km /day/km²	13.4	> 50,000 veh-km /day/km²	16.2
	10,000-25,000 veh-km /day/km²	24.6		
Walking traffic Density	< 100 walk-km/day/km²	18.9	1,000-10,000 walk-km/day/km²	35.4
	100-1,000 walk-km/day/km²	28.1	> 10,000 walk-km/day/km²	17.6

Urban

Rural

Urban branch:

Road section
- Road user going straight
- Road user without duty to give way
- Speed limit 50 km/h

Pedestrian crossing
- Minor local road

Good public transport accessibility
- Hidden pedestrian
- High-rise residential area
- Medium population density
- Medium-high vehicle traffic
- Medium walking traffic
- **C1**

Pedestrian more than 60 y.o.
- Low HH index
- Intermediate or rural municipality
- Residential area
- Low population density
- Low vehicle traffic
- Low-medium walking traffic
- **C2**

Pedestrian less than 19 y.o.
- Pedestrian more than 60 y.o.
- Rural or peripheral municipality
- Jutland
- Shopping and residential area
- Low population density
- Low vehicle traffic
- Low walking traffic
- **C3**

Intoxicated road user

Pedestrian crossing
- Hidden pedestrian
- Bicycle
- Road user 20-39 y.o.
- Two-way divided road
- Major local road
- Several leisure locations
- Copenhagen
- High public transport accessibility
- Shopping and residential area
- High population density
- High vehicle traffic
- High walking traffic
- **C4**

Pedestrian on the road
- Moped or motorcycle
- Road user less than 19 y.o.
- Male road user
- Road user without duty to give way
- Medium/minor local road
- Intermediate or rural municipality
- Good public transport accessibility
- Residential area
- Low-medium population density
- Medium vehicle traffic
- Low-medium walking traffic
- **C5**

Intersection
- Road user turning
- Pedestrian crossing
- Sober road user
- Major local road
- Darkness illuminated
- High public transport accessibility

Pedestrian 20-39 y.o.
- Car or bus
- Male road user
- Several leisure locations
- Urban municipality
- Copenhagen
- Shopping and residential area
- High population density
- High vehicle traffic
- High walking traffic
- **C6**

Female road user
- Road user with green light
- Road user at stop sign
- 4-legged intersection
- Bad weather
- Wet road surface
- Poor visibility
- Two-way divided road
- Low HH index
- Medium-high population density
- Medium vehicle traffic
- Medium walking traffic
- **C7**

Rural branch:

Road section
- Driver going straight
- Pedestrian on the road
- Car or heavy vehicle
- National or regional road
- Speed limit over 80 km/h
- Male pedestrian
- Intoxicated pedestrian
- Weekend
- Evening and night
- Darkness not illuminated
- Poor visibility
- Rural or peripheral municipality
- High HH index
- Jutland
- Low public transport accessibility
- Low population density
- Low vehicle traffic
- Low walking traffic
- **C8**

Figure 1. Visualisation of cluster characteristics for fully identified records
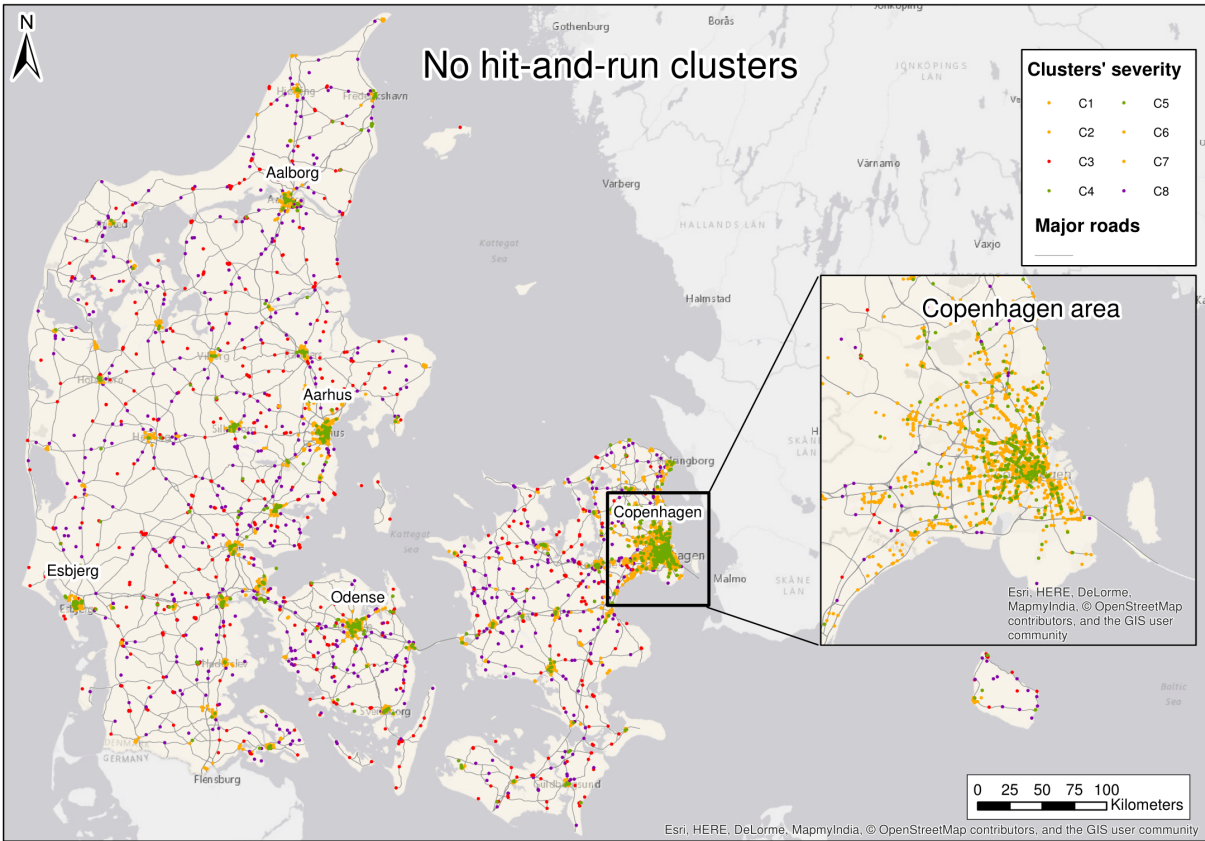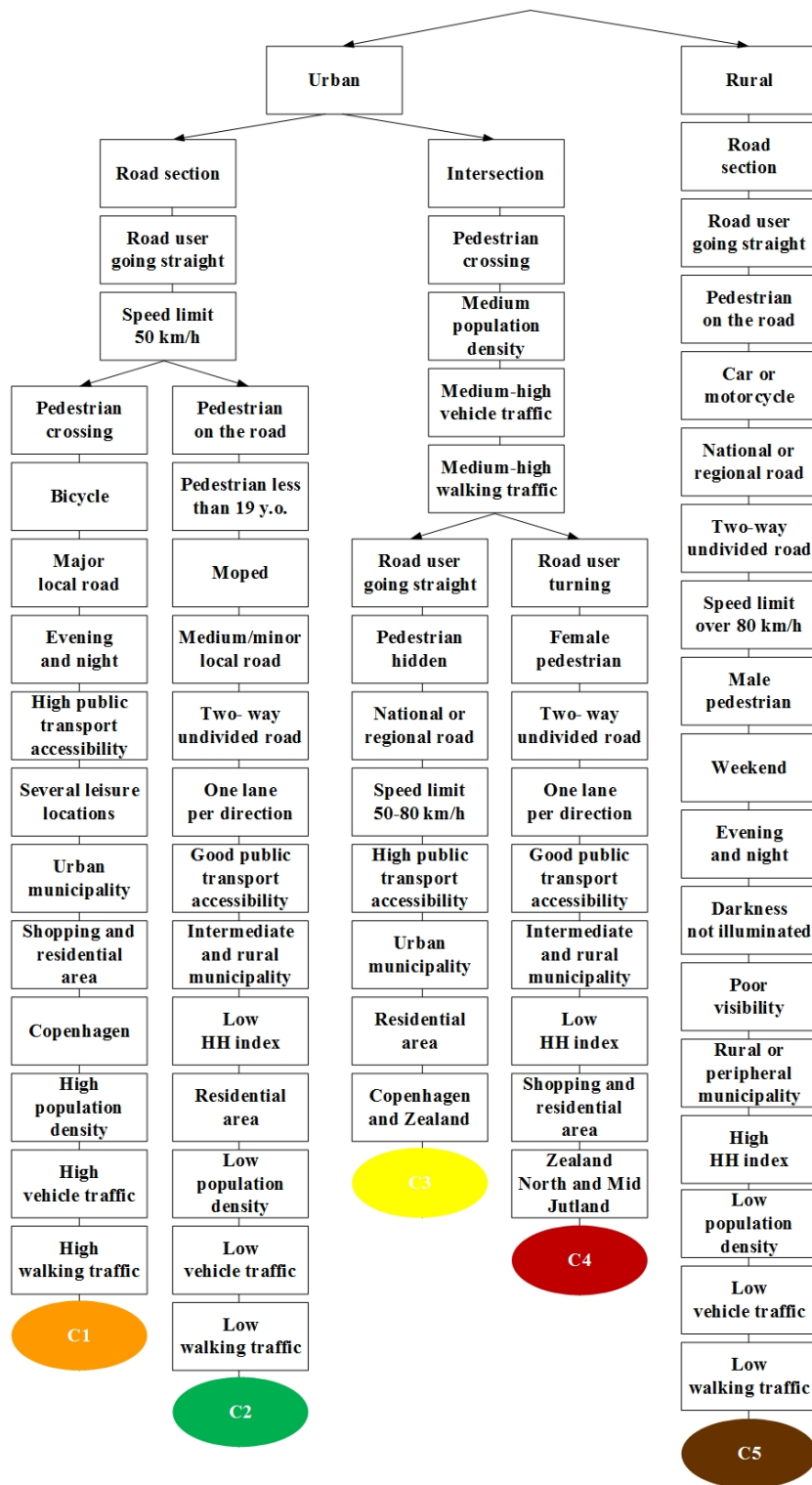
Figure 2. Map of fully identified records

Figure 3. Visualisation of cluster characteristics for hit-and-run records
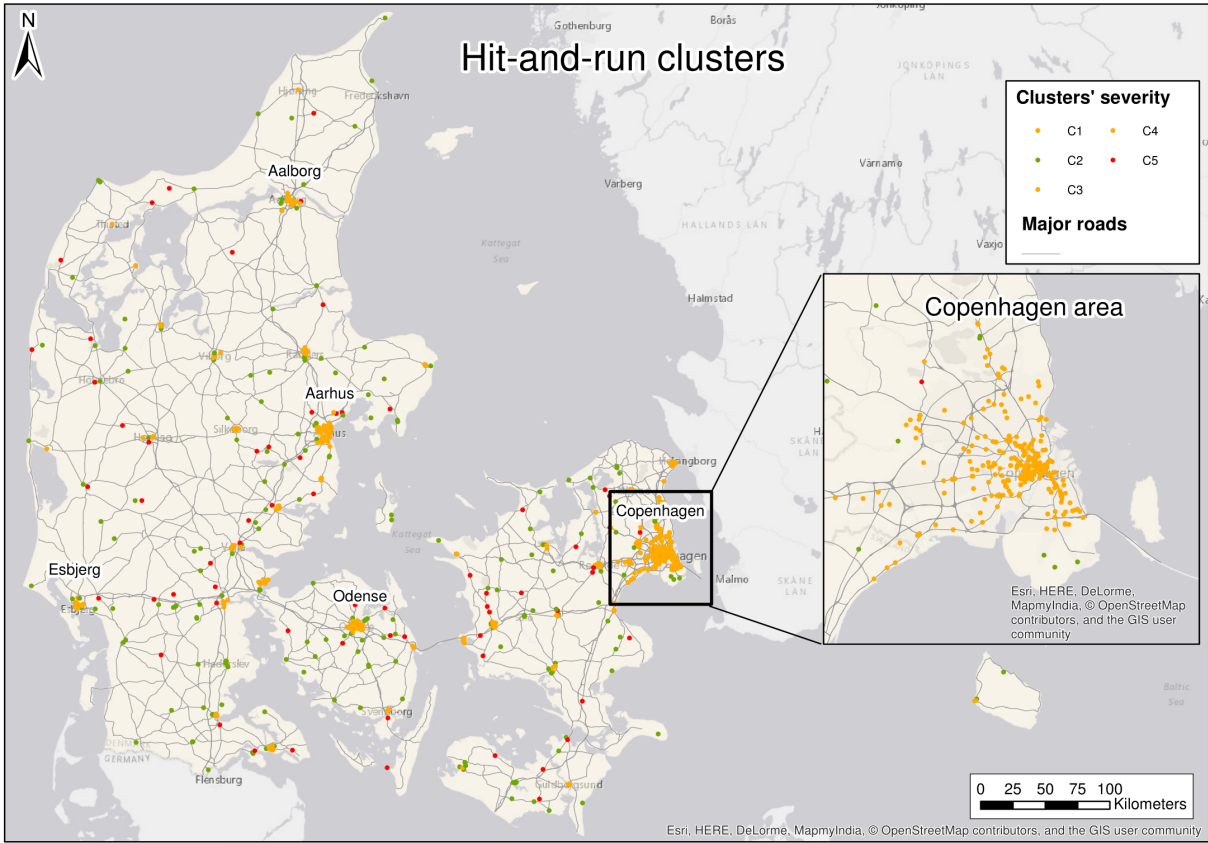
Figure 4. Map of hit-and-run records

Appendix A1: LCA results for the fully identified records

Although all the variables from the joint crash database were used, table A1 presents the results of the LCA for the 8 cluster solution in terms of distribution of the characteristics for the variables that had a significant difference with respect to their distribution in the sample for at least one of the clusters. The table presents also the probability of a crash belonging to each cluster that varies between 5.4% and 19.9%.

Table A1. LCA results for the fully identified sample

Variable	Category	Sample	C1	C2	C3	C4	C5	C6	C7	C8
		100.0	19.9	14.6	9.6	5.4	6.7	17.9	15.5	10.4
Characteristics of the pedestrian										
Gender	Male	46.5	50.6	45.9	49.5	37.9	42.3	46.7	32.9	64.4
	Female	53.5	49.4	54.1	50.5	62.1	57.7	53.3	67.1	35.6
Age	<= 9 years old	7.5	9.4	10.7	15.3	7.2	9.2	3.4	2.8	5.5
	10-14 years old	7.9	9.5	10.6	11.8	2.9	9.7	4.4	6.3	7.5
	15-19 years old	10.8	12.1	11.4	11.3	2.8	12.9	9.1	9.8	14.5
	20-29 years old	15.2	12.7	8.7	7.3	17.4	9.8	25.7	18.4	16.2
	30-39 years old	10.4	9.1	8.8	5.0	14.6	10.2	14.7	8.9	12.6
	40-49 years old	11.1	9.8	9.0	6.2	11.8	12.4	12.5	12.8	14.7
	50-59 years old	10.3	9.1	9.6	9.2	10.2	9.5	9.1	12.9	12.8
	60-69 years old	9.5	7.2	10.0	9.3	15.9	10.7	8.4	12.0	7.1
	>= 70 years old	17.3	21.1	21.2	24.6	17.2	15.6	12.7	16.1	9.1
Intoxicated	No	89.1	87.6	91.9	92.1	95.2	95.2	82.1	95.0	80.9
	Yes	10.9	12.4	8.1	7.9	4.8	4.8	17.9	5.0	19.1
Behaviour	Crossing from the right	36.5	32.3	33.6	28.4	55.0	28.7	47.9	48.5	14.6
	Crossing from the left	27.4	24.4	31.6	23.8	12.6	15.9	23.8	51.1	16.7
	Hidden	8.8	12.9	7.8	8.7	17.0	11.8	9.8	0.4	7.5
	Walking along the road	5.6	2.5	3.1	5.6	2.5	9.4	1.5	0.0	29.2
	Walking on the pavement	5.0	5.6	4.9	7.2	8.0	17.2	3.7	0.0	2.6
	Standing on the road	10.6	12.4	10.0	12.3	4.9	17.1	6.9	0.0	27.0
	Hit by reversing vehicle	6.1	9.9	9.0	14.0	0.0	-0.1	6.4	0.0	2.4
Characteristics of the other road user										
Vehicle	Bicycle	6.5	0.0	0.0	0.0	73.0	35.3	0.0	0.0	2.2
	Moped/motorcycle	7.2	1.5	1.0	3.5	21.1	56.0	2.4	0.5	9.3
	Car	74.3	87.3	87.5	83.1	5.9	7.5	79.5	89.1	70.2
	Van	4.9	4.8	5.7	7.3	0.0	0.9	4.8	4.7	7.2
	Heavy vehicle	2.6	2.0	2.4	3.0	0.0	0.0	1.8	2.5	8.5
	Bus	4.5	4.4	3.4	3.1	0.0	0.3	11.5	3.2	2.6
Gender	Male	73.1	71.5	66.9	67.4	69.4	84.8	82.5	67.0	77.4
	Female	26.9	28.5	33.1	32.6	30.6	15.2	17.5	33.0	22.6
Age	<= 17 years old	2.8	0.0	0.0	2.8	5.1	27.8	0.7	0.0	2.9
	18-19 years old	5.5	5.4	5.6	6.8	3.9	11.5	4.1	4.3	5.9
	20-24 years old	11.0	10.3	9.7	7.9	20.0	8.6	12.5	10.4	12.1
	25-29 years old	8.9	9.2	6.8	6.8	13.3	5.4	11.7	9.3	8.1
	30-39 years old	18.5	18.5	16.4	15.1	22.1	12.7	23.6	18.6	17.9
	40-49 years old	19.1	19.9	20.3	21.2	14.2	12.4	19.1	18.3	21.8
	50-59 years old	16.3	16.5	19.3	17.7	14.5	10.7	16.5	16.8	13.5
	60-69 years old	10.5	12.2	11.5	11.7	5.2	7.2	8.9	11.7	10.9
	>= 70 years old	7.4	8.0	10.4	10.0	1.7	3.7	2.9	10.6	6.9
Intoxicated	No	96.8	98.5	98.1	96.1	90.9	87.7	98.3	99.0	95.7
	Yes	3.2	1.5	1.9	3.9	9.1	12.3	1.7	1.0	4.3
Driving license	No license required	9.9	0.0	0.0	0.0	89.9	65.3	0.0	0.0	6.6
	Valid license	86.2	98.7	99.5	95.6	0.0	15.0	98.0	97.0	89.4
	Invalid license	3.9	1.3	0.5	4.4	10.1	19.7	2.0	3.0	4.0

Behaviour	Straight	70.3	84.4	73.6	70.6	90.0	80.7	66.9	25.1	94.5
	Turn right	5.3	0.0	4.5	2.8	0.8	1.8	7.3	18.8	0.3
	Turn left	13.3	0.0	8.2	5.4	1.3	0.4	15.6	56.1	0.1
	Reversing	6.0	10.0	8.9	14.1	0.0	0.0	6.5	0.0	2.4
	Jump off the road	5.1	5.6	4.8	7.1	7.9	17.1	3.7	0.0	2.7
Traffic control	Traffic light: green	15.2	2.1	5.7	3.2	6.2	1.3	25.5	55.9	0.0
	Traffic light: red	2.5	1.0	1.1	0.0	15.3	3.5	3.8	2.9	0.0
	Stop or give way sign	13.5	8.6	16.3	12.6	3.7	5.0	14.1	32.5	1.0
	No duty to give way	68.8	88.3	76.9	84.2	74.8	90.2	56.6	8.7	99.0

Characteristics of the crash

Type of day	Weekday	79.1	79.9	80.8	75.5	85.6	81.6	74.6	85.9	70.9
	Weekend	20.9	20.1	19.2	24.5	14.4	18.4	25.4	14.1	29.1
Light	Daylight	61.7	67.7	67.6	66.7	76.7	75.0	61.2	48.5	41.7
	Twilight	4.9	3.8	6.5	6.2	4.1	6.7	3.7	5.8	4.0
	Darkness	6.6	1.7	1.3	2.9	0.0	4.0	0.5	1.1	50.2
	Illumination	26.8	26.8	24.6	24.2	19.2	14.3	34.6	44.6	4.1
Visibility	Good	92.8	95.5	93.8	93.7	96.6	96.0	94.0	87.9	87.0
	Poor	7.2	4.5	6.2	6.3	3.4	4.0	6.0	12.1	13.0
Weather	Good	81.0	84.1	81.9	84.0	90.4	89.7	82.4	65.0	81.5
	Bad	19.0	15.9	18.1	16.0	9.6	10.3	17.6	35.0	18.5
Road surface	Dry	66.3	69.8	65.8	69.4	77.4	79.5	70.1	46.9	65.7
	Wet/slippery	33.7	30.2	34.2	30.6	22.6	20.5	29.9	53.1	34.3

Design measures

Road category	Expressway and motorway	1.1	0.0	0.0	0.2	0.0	0.5	0.0	0.4	9.8
	National	12.8	8.5	6.7	10.7	14.0	11.5	13.1	18.6	22.4
	Regional	13.2	9.8	13.2	17.0	9.4	15.3	5.3	16.9	24.2
	Major local	37.4	36.8	31.1	16.5	59.2	28.3	59.0	44.6	13.0
	Medium local	15.8	16.5	21.1	26.7	7.6	19.1	7.1	14.7	15.1
	Minor local	19.7	28.4	27.9	28.9	9.8	25.3	15.5	4.8	15.5
Road directions	One-way	5.8	5.4	5.0	1.3	8.1	2.1	12.1	5.4	2.5
	Two-way undivided	81.3	79.8	91.3	97.3	71.0	87.5	74.4	69.3	86.7
	Two-way divided	12.9	14.8	3.7	1.4	20.9	10.4	13.5	25.3	10.8
Location	4-legged intersection	21.3	3.2	10.8	7.9	20.8	6.4	33.3	68.3	2.2
	3-legged intersections	10.7	5.7	12.4	6.8	7.2	4.8	13.4	22.8	4.4
	Roundabout	1.2	0.3	3.2	0.6	0.0	0.9	0.1	3.0	0.2
	Other intersection type	5.3	6.1	8.8	11.3	0.8	3.9	3.5	3.6	1.9
	Straight road	57.3	83.2	63.1	69.8	54.8	61.5	48.8	2.2	83.0
	Curve	1.6	1.3	1.6	3.4	0.0	1.5	0.7	0.0	5.2
	Bicycle lane or path	2.6	0.2	0.1	0.2	16.4	21.0	0.2	0.1	3.1
Speed limit	<= 50 km/h	5.9	6.8	7.4	6.2	7.6	14.0	5.8	2.0	1.9
	50 km/h	74.1	81.2	85.5	90.7	85.7	71.1	89.6	69.7	5.0
	50-80 km/h	10.9	11.6	5.9	2.6	6.4	13.8	4.6	27.2	11.5
	80-130 km/h	9.1	0.4	1.2	0.5	0.3	1.1	0.0	1.1	81.6

Density measures at the street buffer level (within 500m of the crash location)

Roadside land use	Shopping street	17.4	17.3	19.5	22.2	23.1	17.5	25.9	12.0	0.4
	Industrial area	6.4	9.4	8.7	3.6	4.4	7.2	4.6	7.6	1.6
	High residential area	35.6	40.2	24.5	7.9	61.0	24.9	64.7	43.3	0.7
	Low residential area	19.2	19.6	30.3	54.8	4.4	31.6	0.8	18.5	2.6
	Buildings with no access	6.3	7.8	7.3	3.4	4.9	5.5	3.2	12.1	2.7
	Spare buildings	15.1	5.7	9.7	8.1	2.2	13.3	0.8	6.5	92.0
Low speed road (< 30 km/h)	0 %	25.1	15.4	25.4	51.5	4.0	26.8	2.7	12.8	84.9
	0-10 %	36.8	39.6	38.8	27.2	39.9	36.2	46.9	45.4	5.8
	10-25 %	26.4	30.2	22.9	15.4	38.2	26.4	37.1	29.4	4.7
	> 25%	11.7	14.8	12.9	5.9	17.9	10.6	13.3	12.4	4.6

Density measures at the zone level (per NTM zone)

Population Density	< 100 inhabitants/km^2	15.3	1.2	0.0	77.5	0.3	8.3	0.0	0.7	67.0
	100-1,000 inhabitants/km^2	25.1	2.3	95.3	22.5	1.1	42.2	0.0	16.3	29.6
	1,000-5,000 inhabitants/km^2	40.2	90.7	4.7	0.0	40.5	49.2	18.8	79.3	3.4
	> 5,000 inhabitants/km^2	19.4	5.8	0.0	0.0	58.1	0.3	81.2	3.7	0.0

Destination accessibility measures

Leisure locations	0	23.7	13.8	26.3	37.8	0.4	28.7	0.0	10.0	95.2
	1-10	40.3	55.0	52.6	59.8	24.6	52.8	7.1	60.0	4.8
	> 10	36.0	31.2	21.1	2.4	75.0	18.5	92.9	30.0	0.0
Distance to transit measures										
Public transport stops / stations	0	9.1	2.8	4.4	10.6	0.0	4.0	0.1	0.8	62.2
	1-5	59.7	67.1	75.1	86.2	50.3	76.0	33.8	60.2	37.8
	> 5	31.2	30.1	20.5	3.2	49.7	20.0	66.1	39.0	0.0
Diversity measures										
Region	Copenhagen	37.6	41.3	13.1	6.0	71.4	20.5	79.8	40.8	10.3
	Zealand	20.1	21.6	25.9	24.8	10.0	22.4	9.1	22.5	23.9
	Fyn	8.0	4.7	10.2	14.4	5.5	8.0	4.5	5.8	16.0
	North Jutland	12.1	12.1	18.1	21.0	2.0	14.1	0.0	10.6	22.0
	Mid-Jutland	12.8	12.5	22.5	20.5	2.4	21.7	0.0	11.4	16.2
	South Jutland	9.4	7.8	10.2	13.3	8.7	13.3	6.6	8.9	11.6
Area	Urban	87.7	97.7	95.2	97.5	99.0	93.6	99.6	97.1	4.9
	Rural	12.3	2.3	4.8	2.5	1.0	6.4	0.4	2.9	95.1
Rural index	Urban	56.5	66.7	25.0	13.5	95.9	39.3	100.0	66.9	21.9
	Intermediate	14.8	16.7	28.0	17.0	2.7	22.2	0.0	14.8	17.5
	Rural	22.4	16.2	31.7	51.3	1.3	33.0	0.0	18.1	43.2
	Peripheral	6.3	0.4	15.3	18.2	0.1	5.5	0.0	0.2	17.4
HH index	< 0.50	39.8	45.4	51.5	43.5	36.5	42.6	29.0	43.0	23.3
	0.50-0.75	25.6	22.9	25.0	33.1	24.6	27.9	30.8	23.2	17.8
	0.75-0.99	22.4	16.6	15.6	14.2	19.5	18.8	22.4	19.3	15.2
	1.00	12.2	15.1	7.9	9.2	19.4	10.7	17.8	14.5	43.7
Demand measures (per NTM zone)										
Road traffic Density	< 5,000 veh-km/day/km²	23.8	1.3	35.0	91.9	0.0	21.7	0.0	2.5	74.1
	5,000-10,000 veh-km /day/km²	13.2	5.8	47.6	6.7	1.0	20.7	0.7	8.7	14.7
	10,000-25,000 veh-km /day/km²	24.9	54.7	14.6	0.9	12.0	42.8	2.3	45.5	8.0
	25,000-50,000 veh-km /day/km²	21.9	27.0	2.7	0.5	46.4	11.4	44.3	30.3	2.4
	> 50,000 veh-km /day/km²	16.2	11.2	0.1	0.0	40.6	3.4	52.7	13.0	0.8
Walking traffic Density	< 100 walk-km/day/km²	19.4	0.0	6.1	96.5	0.0	14.1	0.0	0.4	78.8
	100-1,000 walk-km/day/km²	27.7	17.5	93.9	3.5	0.0	49.7	0.0	30.3	19.9
	1,000-10,000 walk-km/day/km²	35.9	82.1	0.0	0.0	48.5	36.2	21.1	68.7	1.3
	> 10,000 walk-km/day/km²	17.0	0.4	0.0	0.0	51.5	0.0	78.9	0.6	0.0

Appendix A2: LCA results for the hit-and-run records

Similar to the fully identified sample, table 4 presents the results of the LCA for the 5 cluster solution in terms of distribution of the characteristics for the variables that had a significant difference with respect to their sample distribution for at least one of the clusters. The probability of a crash belonging to each cluster varies between 7.4% and 25.3%.

Table A2. LCA results for the hit-and-run sample

Variable	Category			Percent			
		Sample	C1	C2	C3	C4	C5
		100.0	25.3	22.1	26.8	18.4	7.4
Characteristics of the pedestrian							
Gender	Male	50.1	51.6	53.1	48.2	41.5	64.4
	Female	49.9	48.4	46.9	51.8	58.5	35.6
Age	<= 9 years old	2.5	1.4	2.2	3.9	2.6	1.5
	10-14 years old	6.8	1.3	11.2	10.0	4.0	7.4
	15-19 years old	15.9	11.5	23.2	12.4	18.3	16.1
	20-29 years old	21.7	26.2	16.5	24.4	20.7	14.9
	30-39 years old	11.9	14.0	9.4	13.8	8.8	13.3
	40-49 years old	13.2	12.4	18.7	11.8	10.4	11.8
	50-59 years old	11.2	10.2	8.7	10.1	12.3	23.3
	60-69 years old	7.0	11.5	2.4	3.6	11.2	7.3
	>= 70 years old	9.8	11.5	7.7	10.0	11.7	4.4
Intoxicated	No	91.2	85.2	91.4	95.6	91.5	94.1
	Yes	8.8	14.8	8.6	4.4	8.5	5.9
Behaviour	Crossing from the right	36.6	45.3	25.4	44.7	37.1	8.8
	Crossing from the left	19.6	15.8	15.1	19.4	34.0	10.3
	Hidden	7.1	8.4	5.9	11.6	1.9	2.9
	Walking along the road	8.3	2.6	17.4	2.0	0.0	44.3
	Walking on the pavement	9.8	5.8	13.0	9.7	11.4	10.3
	Standing on the road	13.6	12.8	17.8	10.6	9.9	23.4
	Hit by reversing vehicle	5.0	9.3	5.4	2.0	5.7	0.0
Characteristics of the other road user							
Vehicle	Bicycle	7.0	16.5	1.0	5.3	4.0	5.9
	Moped/motorcyle	12.9	6.9	18.0	14.3	9.8	20.9
	Car	76.6	71.9	78.0	77.3	83.2	68.8
	Van	3.5	4.7	3.0	3.1	3.0	4.4
Behaviour	Straight	71.8	74.5	76.9	70.7	57.2	88.2
	Turn right	6.6	4.2	3.6	8.3	12.9	1.5
	Turn left	6.7	6.1	1.0	9.3	13.0	0.0
	Reversing	5.2	9.3	5.5	2.0	5.6	0.0
	Jump off the road	9.7	5.9	13.0	9.7	11.3	10.3
Characteristics of the crash							
Type of day	Weekday	75.4	68.5	73.8	80.1	82.7	67.9
	Week-end	24.6	31.5	26.2	19.9	17.3	32.1
Light	Daylight	57.7	57.1	57.9	61.1	57.7	47.4
	Twilight	7.1	5.5	8.7	5.4	10.3	5.8
	Darkness	5.4	2.5	4.4	1.2	1.8	42.3
	Illumination	29.8	34.9	29.0	32.3	30.2	4.5
Visibility	Good visibility	93.0	91.3	96.3	94.6	92.2	85.4
	Poor visibility	7.0	8.7	3.7	5.4	7.8	14.6
Weather	Good	86.2	86.4	88.1	89.2	81.0	82.4
	Bad	13.8	13.6	11.9	10.8	19.0	17.6
Road surface	Dry	76.1	80.4	76.5	76.0	71.5	72.0
	Wet/slippery	23.9	19.6	23.5	24.0	28.5	28.0
Design measures							
Road category	Expressway and motorway	0.7	0.0	0.0	0.8	0.0	5.8

	National	11.2	6.8	3.9	17.0	11.5	26.3
	Regional	10.7	4.6	9.5	12.4	11.4	26.3
	Major local	35.8	56.0	26.7	32.4	35.3	7.3
	Medium local	17.4	8.5	28.4	19.0	14.6	16.3
	Minor local	24.2	24.1	31.5	18.4	27.2	18.0
Road directions	One-way	8.2	21.2	3.1	3.4	6.6	0.0
	Two-way undivided	80.7	72.8	95.0	67.0	87.5	97.0
	Two-way divided	11.1	6.0	1.9	29.6	5.9	3.0
Location	4-legged intersection	18.3	22.2	5.8	23.5	26.9	1.5
	3-legged intersections	11.1	10.5	10.3	13.2	13.0	3.0
	Roundabout	2.0	0.0	2.5	0.7	7.1	0.0
	Other intersection type	4.0	2.6	5.9	6.3	1.8	0.0
	Straight road	59.3	61.3	70.3	50.2	47.7	80.9
	Curve	2.4	0.9	4.8	1.2	1.9	5.9
	Bicycle lane or path	2.9	2.5	0.4	4.9	1.6	8.7
Speed limit	<= 50 km/h	6.7	7.2	8.0	6.6	4.1	7.3
	50 km/h	79.4	92.1	87.8	71.2	93.4	4.6
	50-80 km/h	8.5	0.7	3.3	22.2	2.5	16.1
	80-130 km/h	5.4	0.0	0.9	0.0	0.0	72.0

Density measures at the street buffer level (within 500m of the crash location)

Roadside land use	Shopping street	18.7	30.8	19.3	8.3	24.0	0.1
	Industrial area	5.5	2.1	6.9	11.9	1.2	0.0
	High residential area	38.3	63.6	12.2	38.0	50.5	0.1
	Low residential area	19.7	0.0	48.2	24.8	12.8	0.1
	Buildings with no access	7.0	2.1	5.8	11.9	7.2	8.7
	Spare buildings	10.8	1.4	7.6	5.1	4.3	91.0
Low speed road (< 30 km/h)	0 %	24.3	3.4	45.3	15.9	13.5	90.8
	0-10 %	35.7	41.3	24.3	37.1	52.5	3.3
	10-25 %	27.3	41.4	19.3	26.6	27.6	4.5
	> 25%	12.7	13.9	11.1	20.4	6.4	1.4

Density measures at the traffic zone level (per NTM zone)

Population density	< 100 inhabitants/km^2	12.0	0.0	32.5	0.4	0.0	64.6
	100-1,000 inhabitants/km^2	25.8	0.4	64.2	8.7	37.5	30.9
	1,000-5,000 inhabitants/km^2	39.1	16.8	3.2	83.3	62.4	4.5
	> 5,000 inhabitants/km^2	23.1	82.8	0.1	7.6	0.1	0.0

Density accessibility measures

Leisure locations	0	20.9	0.0	36.5	14.9	12.1	90.9
	1-10	38.6	1.3	56.6	68.6	36.5	9.0
	> 10	40.5	98.7	6.9	16.5	51.4	0.1

Distance to transit measures

Public transport stops / stations	0	7.5	0.9	10.7	2.4	1.2	55.7
	1-5	60.0	35.5	82.0	78.8	46.2	44.2
	> 5	32.5	63.6	7.3	18.8	52.6	0.1

Diversity measures

Region	Copenhagen	36.6	77.5	9.0	52.7	2.0	5.8
	Zealand	22.6	11.5	24.4	23.1	33.5	26.3
	Fyn	5.3	3.4	8.4	5.3	2.5	9.0
	North Jutland	13.1	0.0	22.4	7.1	23.5	26.4
	Mid-Jutland	13.4	0.4	23.9	1.0	34.1	20.6
	South Jutland	9.0	7.2	11.9	10.8	4.4	11.9
Area	Urban	90.9	96.2	99.4	99.2	96.5	2.2
	Rural	9.1	3.8	0.6	0.8	3.5	97.8
Rural index	Urban	57.6	100.0	21.4	92.1	5.8	24.9
	Intermediate	16.0	0.0	20.1	6.7	46.1	17.6
	Rural	19.8	0.0	40.8	0.8	40.3	42.9
	Peripheral	6.6	0.0	17.7	0.4	7.8	14.6
HH index	< 0.50	38.6	21.3	45.9	37.5	61.1	23.8
	0.50 - 0.75	24.3	28.8	29.5	23.6	23.7	24.8
	0.75 - 0.99	19.8	27.7	19.2	16.8	7.7	10.4
	1.00	17.3	22.2	5.4	22.1	7.5	41.0

Demand measures (per NTM zone)							
Road traffic	< 5,000 veh/day/km²	21.6	0.0	60.8	0.0	16.8	69.1
Density	5,000-10,000 veh/day/km²	14.5	1.3	29.9	5.5	24.5	21.9
	10,000-25,000 veh/day/km²	22.8	1.1	8.0	48.2	39.0	8.9
	25,000-50,000 veh/day/km²	23.9	45.5	1.3	31.4	19.7	0.1
	> 50,000 veh/day/km²	17.2	52.1	0.0	14.9	0.0	0.0
Walking traffic	< 100 walk-km/day/km²	15.2	0.0	43.7	0.0	0.0	74.9
Density	100-1,000 walk-km/day/km²	30.9	0.0	56.3	29.8	47.3	23.5
	1,000-10,000 walk-km/day/km²	32.2	14.3	0.0	69.8	52.7	1.6
	> 10,000 walk-km/day/km²	21.7	85.7	0.0	0.4	0.0	0.0