

XV Conference on Transport Engineering, CIT2023

Analyzing pedestrians' crash injury risk factors in Barcelona

Ahmad Aiash^{a*} Francesc Robusté^a

^a Civil Engineering School, UPC – BarcelonaTech, Jordi Girona 1-3, 08034 Barcelona (Spain)

Abstract

Pedestrian crashes can be affected by different factors that can increase or decrease their possibilities of occurring and other consequences related to them, including the level of injury. In Barcelona, in particular, and in Spain, in general, several events happened during both the years 2020 and 2021. The first year had the pandemic and different events that affected the number of traffic crashes, similar to the following year, which had the introduction of applying the speed limit of 30 km/h inside the cities. Therefore, this study analyzes several risk factors related to pedestrian crashes during both the 2020 and 2021 years. A Bayesian network model is employed for this reason to determine the impact of those risk factors through these different years. This is done by calculating the conditional probabilities for the risk factors based on the applied model with determining the accuracy of the employed model in predicting the utilized data. The results show that elderly pedestrians are more likely to have severe or fatal injuries. The evening period shows higher probabilities of having pedestrian crash injuries compared to other periods of the day. Passage for pedestrians that are regulated by traffic lights is found to have a high probability of having severe and fatal injuries.

© 2023 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 15th Conference on Transport Engineering

Keywords: Pedestrian, Risk factors, Injury, Bayesian network, Barcelona

* Corresponding author.

E-mail address: ahmad.aiash@upc.edu

1. Introduction

Due to the fact that traffic crashes are still representing a threat to the lives of road users (Aiash & Robusté, 2021a), certain road users suffer more from the consequences of these crashes. These vulnerable road users are more prone to traffic crash injuries compared to other road users. Based on (WHO, 2022), more than half of the 1.3 million fatalities that are resulted from traffic crashes around the world are for vulnerable road users, including pedestrians, cyclists, and motorcyclists. Several factors, on the other hand, can increase the possibility of having more severe injuries that are resulted from these crashes. Risk factors can include the characteristics of the vulnerable road users such as their gender and their age category, the temporal factors represented by the time when the crash occurred, and the spatial characteristics where the crash happened.

Therefore, several studies have focused on this category of road users in order to prevent or lessen the number of traffic crashes that are causing different levels of injuries so that the number of fatalities can be reduced. For age categories, children and elderly road users showed a higher number of fatalities compared to other groups (Chang, 2008) (Ponnaluri & Nagar, 2010). The correlation between the age of the pedestrian and the severity of the crash is also confirmed by another study (HU, et al. 2020). For the gender of the injured pedestrian, males represented the majority of pedestrian fatalities compared to females in the U.S. (Clifton & Livi, 2005). Another study (Sun, et al. 2019) examined several risk factors related to pedestrian crashes by employing the latent class clustering method and multinomial logit regression. The results have shown that pedestrians with involvement in drugs or alcohol were more likely to have severe and fatal injuries, similar to elderly pedestrians who were older than 65 years old and in adverse weather conditions. For other findings of the same study, pedestrian crash fatalities can be highly impacted when crossing a road far from an intersection. Speed limit (Munira, et al. 2020) and the approach speed of the vehicle toward a pedestrian (Chakraborty, et al. 2019) has an impact on increasing the possibility of fatal injuries for pedestrians. Additionally, periods of the day can have a different impact on injury severity (Pahukula, et al. 2015). For socioeconomic status, the low socioeconomic status of the pedestrian was found to increase the odds of having injuries and fatalities (Cubbin & Smith, 2002) (Azetsop, 2010). The characteristics of the area and road can also impact the severity of pedestrian traffic crashes. A study (Zajac & Ivan, 2003) examined the impact of road features and the types of areas in a rural area. The results have shown that clear road width, village, downtown fringe, and low-density residential areas had a higher level of traffic crash severities.

This study attempts to analyze several risk factors that are related to traffic crashes that occurred for pedestrians in the city of Barcelona by utilizing traffic crash data that is provided by the open data service of the city during the years 2020 and 2021. Slight, severe, and fatal injuries are included to be analyzed. A Bayesian network model is then employed to determine the conditional probabilities for the risk factors in order to grasp their impact on the level of injuries that occurred for pedestrians. The remainder of the paper is organized as follows: the methodology section is set to be section two. Section three presents the results. Lastly, the conclusions are depicted in section four.

2. Methodology

Barcelona's City Hall Open Data Service is the database that is utilized in this study to gather data related to pedestrian crashes that occurred in Barcelona city (Ajuntament de Barcelona's Open Data Service, 2021). In figure 1, the total number of traffic crashes and different levels of injuries are depicted. The consistency of pedestrian crash numbers can be seen through the years from 2011 to 2019. Then, in 2020, the COVID-19 pandemic started with a variety of restrictions that limited the movement of road users that, resulted in a drastic decrease in the number of pedestrian crashes with different consequential levels of injuries. However, the decrement stopped in 2021 as the total number of these crashes started to increase again.

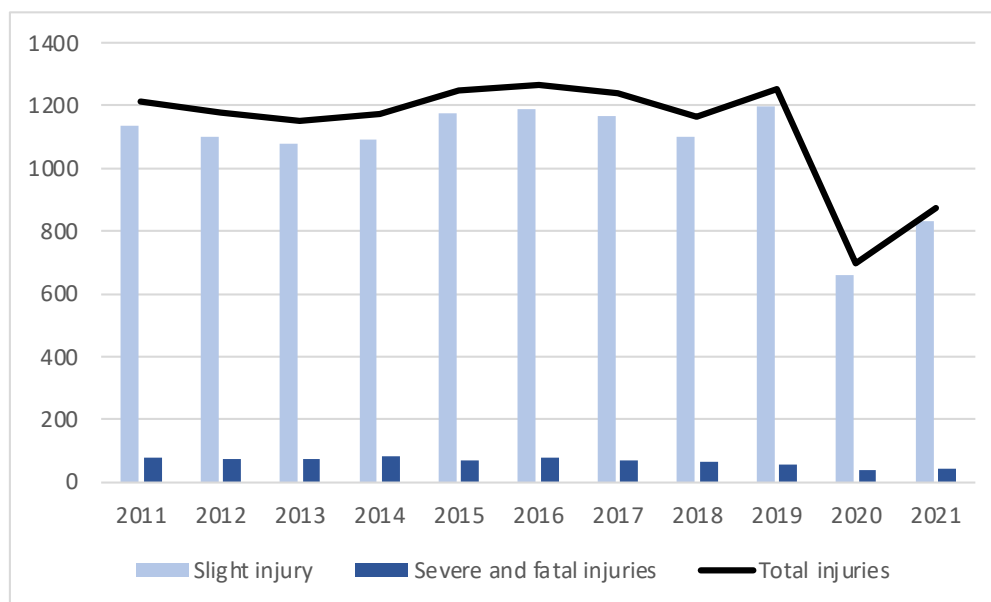


Fig. 1. Pedestrian injuries across 10 years interval in Barcelona

The risk factors that are included to be analyzed with the employed model are shown in table 1 and table 2. It is worth mentioning that all traffic crashes with their risk factors happened during 2020 without merging 2020 year with 2021 due to the differences between the pandemic year as many events occurred in that year when the movement restrictions were imposed because of the pandemic. The daytime variable consists of three classes: morning, evening, and night timing, based on the database. The morning period is set to be from 6:00 to 13:00. The evening period is set to be from 14:00 to 21:00. The night period is set to be from 22:00 to 5:00. The gender includes male and female. The age category of the injured person includes the 0-15 group, 16-24 group, 25-40 group, 41- 64 group, and lastly, 65 and older age group for the pedestrian who got injuries resulting from the traffic crash. The last risk factor is the crash site characteristics, including six different sites: a crash on the sidewalk/platform, a crash on the passage that is regulated by the traffic light with having both horizontal and vertical signs, a crash on a passage with having horizontal signs but without having vertical signs, a crash on a passage without traffic lights but with having other vertical and horizontal signs, and last but not least, a crash outside the designated passage area for pedestrian.

Table 1. 2020 pedestrian crash risk factors descriptive data during 2020

Risk factor	Risk factor class	Slight injury	Severe or fatal injury
Day time	Morning	255	8
	Evening	323	25
	Night	37	2
Gender	Male	344	13
	Female	271	22
Age	0-15	91	1
	16-24	56	2
	25-40	112	6
	41-64	210	14
	65+	146	12
Site of crash characteristics	On the sidewalk / Platform	78	1
	Passage regulated by traffic lights	271	18
	In a pedestrian area	28	0
	Passage without vertical signs	10	1
	Passage without traffic lights	86	4
	Outside the designated passage area	142	11

Table 2 – 2021 pedestrian crash risk factors descriptive data during 2020

		Slight injury	Severe or fatal injury
Day time	Morning	315	15
	Evening	411	22
	Night	27	4
Gender	Male	438	21
	Female	315	20
Age	0-15	98	9
	16-24	76	5
	25-40	161	8
	41-64	236	8
	65+	182	11
Site of crash characteristics	On the sidewalk / Platform	98	4
	Passage regulated by traffic lights	317	14
	In a pedestrian area	53	0
	Passage without vertical signs	7	1
	Passage without traffic lights	91	5
	Outside the designated passage area	187	17

2.2. Bayesian network

In this study, a Bayesian network (BN) is employed to determine the conditional probabilities between the risk factors and the dependent variable represented by the injury level of the pedestrian. For the given random variables, BN can provide laconic descriptions for the joint distribution of probability as it is considered a member of probabilistic graphical models. In this study and by following previous proceedings (AIASH & ROBUSTÉ, 2021b) in employing BN, Tree Augmented Naïve Bayes is employed using the IBM Watson Studio software platform with utilizing SPSS modeler to classify the correlation between the four predictors and the level of injury that has happened to the pedestrians that are involved in traffic crashes. The training and testing set size partitions for the 2020 and 2021 years are set to be 70% and 30%, respectively. The classifier of Tree Augmented Naïve Bayes that is employed when utilizing IBM Watson Studio software with utilizing SPSS modeler is based on this book (Friedman, Geiger, & Goldszmidt, 1997). The conditional probabilities are determined as follows:

$$Pr(Y_i | X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j) = \frac{Pr(Y_i)(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j | Y_i)}{Pr(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j)} \propto Pr(Y_i) \prod_{k=1}^n Pr(X_k = x_k^j | \pi_k^j, Y_i) \quad (1)$$

For $d_j = (x_1^j, x_2^j, \dots, x_n^j)$, the data set is d . The case is d_j . This case is classified to which it belongs to i^{th} dependent category. In this study, the target or dependent variable Y_i is the level of injury that consists of two categories including slight injury and severe or fatal injury. The predictor is x . The number of predictors is n . In this study, n is four. The number of non-redundant parameters is K . The parent set of the independent variable is π_k . The conditional probability for each node is represented by $Pr(X_k = x_k^j | \pi_k^j, Y_i)$. In this study, there are four nodes for predictors.

3. Results and discussions

3.1. BN results for 2020 year

For the year 2020, the employed BN is shown in figure 2. The training and testing accuracy for the employed BN is 95.08% and 94.09%, respectively. All of the risk factors are included in the BN by providing their importance

values. The most important risk factor based on this employed BN is the site of crash characteristics, while the least important risk factor is the age with its different included categories.

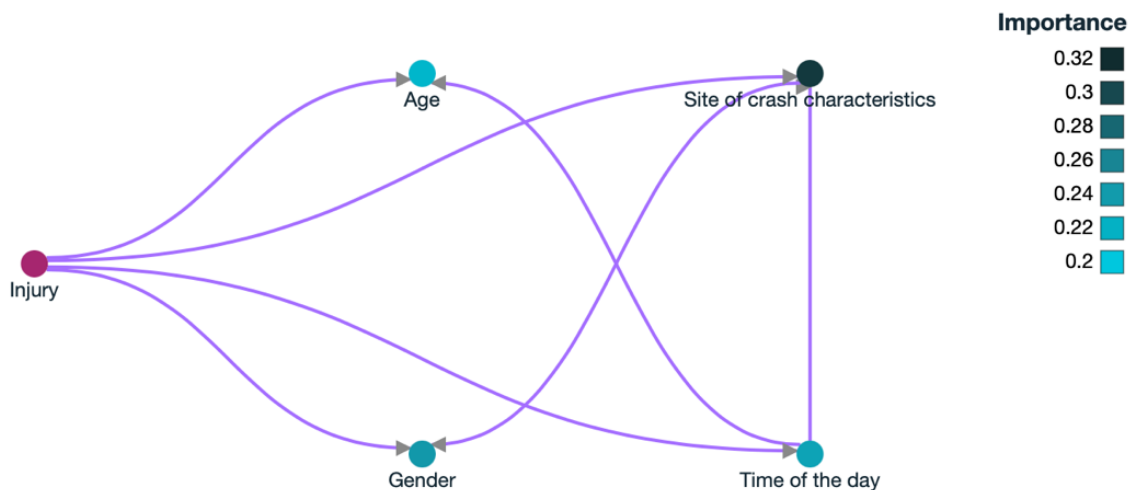


Fig. 2. The structure of the applied BN for 2020

Tables 3, 4, 5, and 6 show the determined conditional probabilities for the chosen risk factors with respect to the dependent variable, that is, the level of injury. Table 3 shows the results for the age factor with two parents’ nodes, including the time of the day and the injury level. For the morning period, severe and fatal injuries are more likely to occur to pedestrians who have an age between 41 years old and older compared to other categories. For slight injuries during the morning period, pedestrians who are 65 years old and older are more likely to have this type of injury compared to other groups. For the evening period, the 65 and older age group is found to have higher odds of having severe and fatal injuries. For slight injuries, the age group between 41 and 64 years old is found to have the highest probability. For the night period, younger age groups are more involved in pedestrian crashes. For severe and fatal injuries, the age group between 25 and 40 years old is more likely to have this type of injury compared to other groups. The age group 41 and 64 years old are more likely to have slight injuries during the night timing period.

Table 3. Age conditional probabilities based on the applied BN for 2020

Parents		Age				
Time of the day	Injury	0-15	16-24	25-40	41-64	65+
Morning	Severe or fatal injury	0.000	0.000	0.000	0.500	0.500
Morning	Slight injury	0.131	0.057	0.159	0.324	0.330
Evening	Severe or fatal injury	0.059	0.118	0.176	0.294	0.353
Evening	Slight injury	0.185	0.104	0.180	0.324	0.207
Night	Severe or fatal injury	0.000	0.000	1.000	0.000	0.000
Night	Slight injury	0.120	0.320	0.200	0.360	0.000

Table 4 shows the results for gender as a risk factor in 2020 year. The parents’ nodes for this risk factor are the site crash characteristics and injury. Males are found to have higher probabilities of having all types of injuries in case of a traffic crash on the sidewalk compared to females. For the crashes that are happening on a passage regulated by traffic lights, females are more likely to have severe and fatal injuries there compared to males, while males are more likely to have slight injuries. Females are found to have higher odds of having slight injuries inside pedestrian areas. For passage without vertical signs, females are more likely to have severe or fatal injuries compared to males, who are more likely to have slight injuries. For passage without traffic lights, males are found to have higher conditional

probabilities for all types of injuries. An interesting finding, females are found to be more involved in pedestrian traffic crashes outside the designated passage area.

Table 4. Gender conditional probabilities based on the applied BN for 2020

Site of crash characteristics	Parents Injury	Gender	
		Female	Male
On the sidewalk / Platform	Severe or fatal injury	0.000	1.000
On the sidewalk / Platform	Slight injury	0.283	0.717
Passage regulated by traffic lights	Severe or fatal injury	0.667	0.333
Passage regulated by traffic lights	Slight injury	0.400	0.600
In a pedestrian area	Slight injury	0.524	0.476
Passage without vertical signs	Severe or fatal injury	1.000	0.000
Passage without vertical signs	Slight injury	0.400	0.600
Passage without traffic lights	Severe or fatal injury	0.500	0.500
Passage without traffic lights	Slight injury	0.456	0.544
Outside the designated passage area	Severe or fatal injury	0.750	0.250
Outside the designated passage area	Slight injury	0.569	0.431

Table 5 presents the results for the site of crash characteristics risk factor with two parents’ nodes: the time of the day and the injury level. Surprisingly, passages regulated by traffic lights show the highest odds for having all types of injuries that are resulted from traffic crashes during different day timings except for night timing as on the sidewalk has the highest odds for having severe or fatal injuries. Table 6 shows the time of day as a risk factor with only one parent node, that is, the injury level. The evening period has the highest odds of having all types of injuries from traffic crashes for pedestrians compared to other day timings.

Table 5. Site of crash characteristics conditional probabilities based on the applied BN for 2020

Time of the day	Parents Injury	Site of crash characteristics					
		On the sidewalk / Platform	Passage regulated by traffic lights	In a pedestrian area	Passage without vertical signs	Passage without traffic lights	Outside the designated passage area
Morning	Severe or fatal injury	0.000	0.500	0.000	0.000	0.167	0.333
Morning	Slight injury	0.085	0.426	0.051	0.023	0.131	0.284
Evening	Severe or fatal injury	0.000	0.529	0.000	0.059	0.059	0.353
Evening	Slight injury	0.167	0.446	0.054	0.005	0.122	0.207
Night	Severe or fatal injury	1.000	0.000	0.000	0.000	0.000	0.000
Night	Slight injury	0.040	0.440	0.000	0.000	0.280	0.240

Table 6. Time of the day conditional probabilities based on the applied BN for 2020

Parents Injury	Time of the day		
	Morning	Evening	Night
Severe or fatal injury	0.250	0.708	0.042
Slight injury	0.416	0.525	0.059

3.2. BN results for 2021

Similar to the 2020 year, the structure of BN for the 2021 year is depicted and shown in figure 3. The prediction accuracy for the training and testing sets are 95.32% and 93.70%, respectively. The structure is also based on the same risk factors as explained before. Eventually, four nodes represent those risk factors, and the fifth node is the injury level which is the dependent variable. Based on the applied BN, the most important risk factor is the age of the injured pedestrian, while the least important variable is gender.

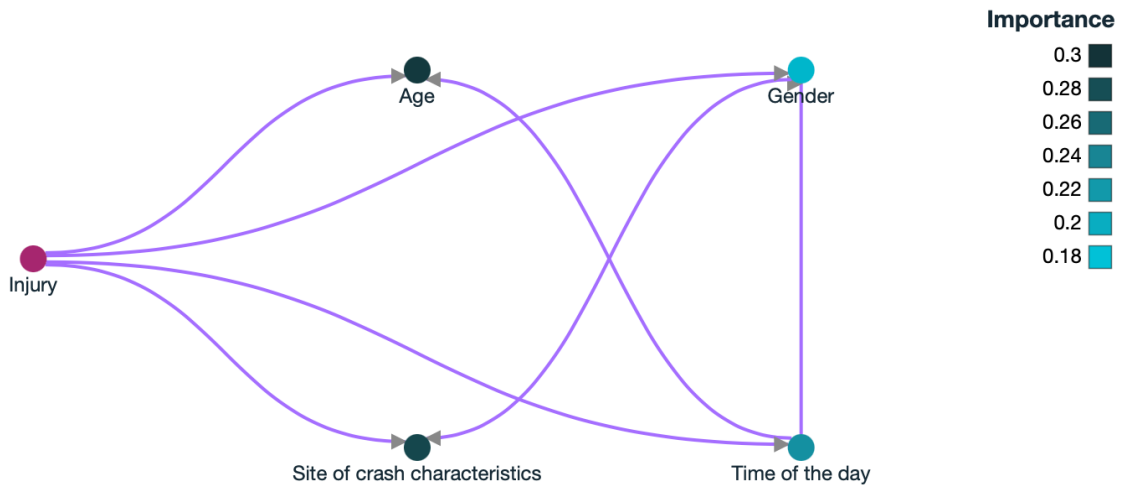


Fig. 3. The structure of the applied BN for 2021 year

Tables 7, 8, 9, and 10 provide the results for the determined conditional probabilities for the employed BN for the year 2021. The results for the age category in table 7 show similar results compared to the year 2020 during the morning period, as 41 years old and older pedestrians are more likely to have severe or fatal injuries, 65 years old and older pedestrians are more likely to have slight injuries. While for the evening period, the results show that the age category from 0 to 15 years old is more involved in severe or fatal injuries, while the 41 to 64 years old age category is more involved in slight injuries during the same time of the day. The night timing period results show that 65 years old and older pedestrians can have higher odds of having severe or fatal injuries, while 41 to 64 years old age group pedestrians show higher probabilities of having slight injuries.

Table 7. Age conditional probabilities based on the applied BN for 2021 year

Time of the day	Parents		Age				
	Injury		0-15	16-24	25-40	41-64	65+
Morning	Severe or fatal injury		0.100	0.100	0.200	0.300	0.300
Morning	Slight injury		0.077	0.047	0.189	0.326	0.361
Evening	Severe or fatal injury		0.313	0.063	0.250	0.188	0.188
Evening	Slight injury		0.175	0.124	0.240	0.287	0.175
Night	Severe or fatal injury		0.000	0.000	0.000	0.000	1.000
Night	Slight injury		0.000	0.238	0.238	0.476	0.048

Table 8 provides the conditional probabilities for the gender risk factor with having two parents' nodes: time of the day and injury. Males are showing higher possibilities of having slight, severe, and fatal injuries compared to females during all periods of the day except during the night, as females are more likely to have severe or fatal injuries compared to males.

Table 8. Gender conditional probabilities based on the applied BN for 2021 year

Time of the day	Parents	Gender	
	Injury	Female	Male
Morning	Severe or fatal injury	0.400	0.600
Morning	Slight injury	0.365	0.635
Evening	Severe or fatal injury	0.438	0.563
Evening	Slight injury	0.440	0.560
Night	Severe or fatal injury	1.000	0.000
Night	Slight injury	0.333	0.667

Table 9 presents the conditional probabilities for the site of crash characteristics risk factor with two parents' nodes: gender and injury nodes. Passage regulated by traffic lights has the highest odds for females for all types of injuries. Table 10 depicts the results for the time of the day risk factor with only one parent node, that is, the injury. The evening period shows the highest probability of having all types of injuries for pedestrians compared to other periods of the day, which is similar to the year 2020.

Table 9. Site of crash characteristics conditional probabilities based on the applied BN for 2021 year

Parents		Site of crash characteristics					
Gender	Injury	On the sidewalk / Platform	Passage regulated by traffic lights	In a pedestrian area	Passage without vertical signs	Passage without traffic lights	Outside the designated passage area
Female	Severe or fatal injury	0.083	0.500	0.000	0.000	0.000	0.417
Female	Slight injury	0.141	0.413	0.066	0.014	0.108	0.258
Male	Severe or fatal injury	0.133	0.200	0.000	0.067	0.200	0.400
Male	Slight injury	0.120	0.399	0.066	0.009	0.158	0.247

Table 10. Time of the day conditional probabilities based on the applied BN for 2021 year

Parents	Time of the day		
	Morning	Evening	Night
Injury			
Severe or fatal injury	0.370	0.593	0.037
Slight injury	0.440	0.520	0.040

4. Conclusions

Vulnerable road users show a higher necessity to be protected compared to other road users due to the fact that they are more prone to traffic crash injuries. Part of the vulnerable road users are the pedestrians, that can be involved in traffic crashes with severe or fatal injuries. Therefore, this study examines different risk factors that can contribute to traffic crash injuries related to pedestrians. The data that is utilized is attributed to Barcelona city with implementing two years, including 2020 and 2021 years. Then, BN models are employed to determine the conditional probabilities between these chosen risk factors and the dependent variable, which is the level of injury resulting from traffic crashes, besides presenting the accuracy of the employed models in predicting the utilized data.

The results show similar findings for both years that are included in this study. Elderly pedestrians show higher probabilities of having severe or fatal injuries during the morning period. Passages regulated by traffic lights are showing a higher number of traffic crash injuries compared to other sites of pedestrian crashes. For the period of the day, the evening period is another risk factor for pedestrian traffic crash injuries. Both BN models that are employed for both years show a high prediction accuracy.

References

- Aiash, A., & Robusté, F. (2021a). Traffic accident severity analysis in Barcelona using a binary probit and CHAID tree. *International Journal of Injury Control and Safety Promotion*, 29(2), 256-264.
- Aiash, a., & Robusté, F. (2021b). Working hours and traffic accident injuries: Case study in Barcelona. *Transportation Research Procedia*, 58, 678-682.
- Ajuntament de Barcelona's Open Data Service. (2021). Retrieved from <https://opendata-ajuntament.barcelona.cat/en/>
- Azetsop, J. (2010). Social Justice Approach to Road Safety in Kenya: Addressing the Uneven Distribution of Road Traffic Injuries and Deaths across Population Groups. *Public Health Ethics*, 3(2), 115–127.
- Chakraborty, A., Mukherjee, D., & Mitra, S. (2019). Development of pedestrian crash prediction model for a developing country using artificial neural network. *International Journal of Injury Control and Safety Promotion*, 26(3), 283-293.
- Chang, D. (2008). *National Pedestrian Crash Report (DOT HS 810 968)*. Washington, DC: National Highway Traffic Safety Administration.
- Clifton, K., & Livi, A. (2005). Gender differences in walking behavior, attitudes about walking, and perceptions of the environment in three Maryland communities. *Research on Women's Issues in Transportation, Transportation Research Board*. Washington, DC.
- Cubbin, C., & Smith, G. (2002). Socioeconomic inequalities in injury: critical issues in design and analysis. *Annual Review of Public Health*, 23, 349-75.
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, 29, 131–163.
- Hu, L., Wu, X., Huang, J., Peng, Y., & Liu, W. (2020). Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China. *Safety Science*, 127, 104710.
- Munira, S., Sener, I., & Dai, B. (2020). A Bayesian spatial Poisson-lognormal model to examine pedestrian crash severity at signalized intersections. *Accident Analysis & Prevention*, 144, 105679.
- Pahukula, J., Hernandez, S., & Unnikrishnan, A. (2015). A time of day analysis of crashes involving large trucks in urban areas. *Accident Analysis & Prevention*, 75, 155-163.
- Ponnaluri, R., & Nagar, F. (2010). Road crash risk among vulnerable population groups in Andhra Pradesh, India. *TRB 89th Annual Meeting Compendium of Papers DVD, Transportation Research Board*. Washington, DC.
- Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis & Prevention*, 124, 50-57.
- WHO. (2022). Road traffic injuries. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- Zajac, S., & Ivan, J. (2003). Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention*, 35(3), 369-379.