Exploring factors affecting crash injury severity of children and adolescents by applying association rule learning and the mixed logit model

Abstract

This paper aims to explore the factors affecting crash injury severity of children and adolescents in Greece. For that reason, disaggregate road crash data including 13431 involving children and adolescents from all regions of Greece for the period 2006-2015 were utilized. To identify factors affecting injury severity and also account for unobserved heterogeneity, a mixed logit model is applied. Moreover, the Association Rule (AR) model Eclat, which is an unsupervised machine learning algorithm for data mining, was applied as a supplementary analysis for data exploration purposes. The aim of the AR was to identify frequent patterns in the data, and more specifically which combinations of crash characteristics are most frequently observed in severe/fatal injuries. The Eclat algorithm indicated that the most common prevailing conditions of crash conditions in fatal/severe injuries are: Clear weather-Male-No intersection-PTW/Bicycle- Urban area, Clear weather-Male-Night-No intersection-PTW/Bicycle, Clear weather-Male-Night-PTW/Bicycle-Urban area. However, the Eclat model did not provide strong statistical evidence, hence it is suggested that there should be a joint utilization of statistical models. Afterwards, results from the mixed logit model indicated that night crashes, crashes outside urban areas, run-off-road collisions or collisions with fixed objects as well as crashes involving bicycles or Powered-Two-Wheelers are associated with higher injury severity of children and adolescents. Interestingly, crashes involving pedestrians are associated with lower injury severity than head-on collisions and run-off-road/collisions with fixed objects. Overall, this study contributes to current knowledge, because to the best of our knowledge this is the first study that addresses unobserved heterogeneity and applies AR models when analysing child/adolescent injury severity. The findings of this study provide useful insights and could assist in unveiling crash risk factors and prioritize programs and measures to promote road safety of children and adolescents.

Keywords: Children, adolescents, crash, injury severity, association rules, mixed logit model

1. Introduction and Background

Crash injury severity analysis has received much attention from researchers worldwide, with many types of studies across different countries and sample frames. Great efforts have been devoted to the severity analysis of vulnerable road users in particular such as pedestrians, cyclists and motorcyclists (Moore et al., 2011; Pour-Rouholamin and Zhou, 2016; Theofilatos and Ziakopoulos, 2018; Xie et al., 2018). For instance, Zhai et al. (2019) analysed pedestrian injury severity with real-time weather data and found that high temperatures and increased rainfall were associated with a higher probability of fatal and severe crashes. Shaheed and Gkritza (2014), explored single-vehicle motorcycle crash severity indicating several specific factors such as speeding, run-off-road, riding without a helmet and so on. An effort was also made by a number of studies to simultaneously address the issue of unobserved heterogeneity (Cerwick et al., 2014; Kim et al., 2010, 2017; Shaheed and Gkritza, 2014; Xu et al., 2016).

However, one aspect that is often neglected is the injury severity analysis of children and adolescents. This is considered an important issue worth to investigate, due to the high traffic mortality of children worldwide. According to the World Health Organization (WHO, 2008), road traffic injuries are the leading cause of death in 10–19 year olds globally. In that report is also suggested that even in European Union countries, where rates are not as high, road traffic injuries still account for 20% of childhood injury deaths.

A few studies had a focus on child pedestrian injury analysis (Adesunkanmi et al., 2000; Bennet and Yiannakoulias, 2015; Koopmans et al., 2015). For instance, Bennet and Yiannakoulias (2015), explored motor-vehicle collisions involving child pedestrians walking to school in Hamilton, Ontario, Canada and suggested that child pedestrian injuries at intersections are correlated with the type of intersection control, traffic volume and land use. Another study by Peek-Asa et al. (2010), suggested that rural crashes involving teens were nearly five times more likely to result to a fatal/severe crash than urban teen crashes. Koopmans et al. (2015), analysed age-based urban pedestrian crash characteristics and compared them with car crashes in order to identify crash characteristics associated with injury severity. In that study, overall incidence was higher for child versus adult pedestrians, but the fatality rate for children was lower.

Research on bicycling injury severity of children and adolescents is relatively limited (Hagel et al., 2015) and it is suggested that in bicycle collisions with motor vehicles, male cyclists as well as those not wearing a helmet were found to increase injury severity. Collisions when cycling on a paved surface or cycling to school or work tend to decrease injury risk.

On the other hand, more research has been carried out having a focus on injury severity when children are vehicle occupants (García-España and Durbin, 2008; Koekemoer et al., 2017; Olsen et al., 2010; Skjerven-Martinsen et al., 2014; Zaveri et al., 2009). There are several studies that have correlated child occupant injury severity with restraint use or misuse, seating position of children (Arbogast et al., 2005; Brown et al., 2010, 2006; Decina and Lococo, 2005; Durbin et al., 2005; Viano et al., 2008). Other common risk factors also include (interestingly) driver restraint use, light conditions and age as well as pedestrian collisions with bicycles. Furthermore, according to García-España and Durbin (2008), head injuries were the most common types of injury (60%), while 21% of 8–12 year old children either did not use the shoulder portion of the car seat belt or did incorrectly place it (behind their back or under their arm). Another general remark is the fact that obesity was found to be highly correlated with paediatric injury severity and distribution of injuries or restraint use (Zaveri et al., 2009).

A recent study by Duddu et al. (2019), investigated risk factors associated with injury severity of teen drivers by applying a partial proportionality odds model. However, the issue of unobserved heterogeneity is not accounted for. The authors found that sports utility vehicles and pickup trucks are more likely to be associated with severely injured teens (compared to passenger cars). Furthermore, teen drivers are more likely to be severely injured on weekdays and peak hours.

Overall, there is a limited number of studies which focused on analysing the impact of various risk factors on child and adolescent injury severity by applying linear statistical models, since chi square tests, relative risk ratios or other descriptive analyses are usually applied in relevant past literature in the field. At the same time, in the few cases where statistical modelling were developed, the common approach was to apply fixed effects binary logistic models. Hence, the issue of unobserved heterogeneity has not been accounted for. Mannering et al. (2016) emphasize the strong need to account for unobserved heterogeneity when analysing crash injury severity data. Similarly, the applicability of Machine Learning (ML) techniques and in particular the Association Rule (AR) mining, has not been used when analysing children and adolescent crashes. To the best of our knowledge, only a few studies have applied AR models in the field of road safety (Das et al., 2019). More specifically, in that particular study (Das et al., 2019), the authors used the apriori method to identify the crash and geometric features which contribute to hit-and-run crashes.

Consequently, our study examines this important issue and aims to address this gap and add to current knowledge by analysing child and adolescent injury severity of all road crashes occurred in Greece for the years 2006 to 2015. It is noted that low- and middle-income countries have received less focus, although these countries account for 93% of child road traffic deaths (WHO, 2008). To account for potential unobserved heterogeneity both on a temporal and a spatial degree, a mixed logit model in which the constant term follows a normal distribution across all Greek counties is applied, while at the same time the variable Year is included in the independent variable list. It is also noted that the Association Rule (AR) algorithm Eclat, which is an unsupervised machine learning algorithm for data mining was applied as a preliminary analysis for data exploration purpose. The aim of the AR is to identify frequent patterns in the data, and most specifically which combinations of crash characteristics are more frequently observed when severe and fatal injuries of children and adolescents occur.

2. Data

The required crash data were collected from the Greek crash database SANTRA, which is provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens. The SANTRA database consists of all crash data in Greece, which are filled in high detail by the Police immediately after the occurrence of a crash. It is noted that disaggregate data require specific treatment because they are confidential and must be used only for a scientific and research purposes.

In our analysis, the crash dataset of interest involved all injured persons under 18 years old (children and adolescents) who were involved in crash in all Greek regions (counties) for a 10 year period, 2006-2015 (13431 children and adolescents in total). The response variable is crash injury severity, which is a binary variable in nature. Two severity levels existed: 0=Slight Injury (SI) and 1=Killed/Severe Injury (KSI). The aim is to investigate factors affecting the outcome of a child/adolescent injury (KSI versus SI), given that the crash has occurred. Hence, we followed the traditional approach to investigate injury severity; that is a comparison between KSI and SI. The comparison with adult injuries was out of the scope of this paper, therefore no injury adult data were collected. Furthermore, in the Greek crash database, fatalities are defined as all persons killed within 30 days after the crash, severe injuries are all persons hospitalized for at least 24 hours, whilst slight injuries are all other persons reported by the police as injured without hospitalization. Other individual and crash specific variables considered were gender, weather, area type, location, vehicle type, illumination and crash type. The year of the crash was also considered in order to explore the temporal evolution of injuries.

3. Analysis Methods

3.1 Association Rule

Most Machine Learning (ML) models focus on numerical data and normally serve predictive purposes. One the other hand, the Association Rule (AR) is perfect for categorical data, is a rule-based ML method which is able to unveil relationships between variables and also identifies strong rules in the basis of interestingness measures.

Suppose that I= $\{i_1, i_2, \dots, i_n\}$ is a set of binary attributes (items) and D= $\{t_1, t_2, \dots, t_n\}$ is a set of transactions (dataset). Each transaction in D contains a subset of items in I. A rule is defined as an implication of the form $X \rightarrow Y,$ where X, Y are itemsets and $X, Y \in I$. The term Support is very important in AR models because it is an indication of how frequently the itemset appears in the dataset.

The most common AR methods are the Apriori and the Eclat methods, which have been initially applied in market basket analysis, in order to identify frequent combination of products that consumers buy together. In our study, we apply the Eclat method which is considered suitable for categorical variables. We base our rules in identifying the most common combination of 3 and 5 variables (respectively) in severe and fatal injuries with a minimum support of 0.1, i.e. to be present in at least 10% of cases. e rules are generated by using 'arules' package of the open source R Software (Hahsler et al., 2005).

3.2 Mixed logit model

To account for unobserved heterogeneity, a random parameters binary logistic model (mixed logit) was applied. In random parameters models, parameters may vary across observations, in a sense that they follow a distribution, e.g. normal, uniform and so on. Following Washington et al. (2011) the random-parameter model has for observation n, outcome probabilities defined as $P_n^m(i)$:

$$P_n^m(i) = \int_x P_n(i) f(\beta|\varphi) d\beta \qquad \text{Eq. (1)}$$

It is noted that $P_n(i)$ is the probability of observation n having a discrete outcome i, whilst f(β|φ) is the density function of β and φ refers to a vector of parameters of the density function (e.g. mean and variance). Lastly:

$$P_n^m(i) = \int_x \frac{\exp[\beta_i X_{in}]}{\Sigma_I \exp[\beta_i X_{in}]} f(\beta|\varphi) d\beta \qquad \text{Eq. (2)}$$

In our approach, a random parameters binary logistic model (mixed logit model) in which the constant term is following a normal distribution is applied and its goodness-of-fit is compared with the fixed effects model. It is noted that the goodness-of-fit of the models can be assessed with the McFadden R^2, which is based on the likelihood ratios of the full model (L_f) and the empty model (L_0). Model comparisons could also be performed via the Akaike Information Criterion (AIC).

4. Results

4.1 Data exploration

A preliminary analysis was firstly carried out. Overall, the total number of injured children and adolescents involved in road crashes in the 51 counties of Greece was 13431, out of which 11702 were classified as slightly injured (SI) and 1729 as killed/severely injured (KSI). Table 1 provides an illustration of the variables considered, potential values as well as indicative descriptive statistics. Figure 1 illustrates the temporal evolution of crash severity by year, showing a steadily decreasing number of total injuries. Figure 2 illustrates the distribution of the number of injuries by severity level for Attica and Thessaloniki and Figure 3 for the rest of Greek counties. It was expected that the majority of injuries would occur in Attica and Thessaloniki regions, because they include the cities of Athens and Thessaloniki respectively (about 50% of the total Greek population).

***** Please insert Table 1 about here *****

***** Please insert Figure 1 about here *****

***** Please insert Figure 2 about here *****

***** Please insert Figure 3 about here *****

To supplement our analysis an Association Rule (AR) model was then applied. When interpreting the results of the AR model regarding the severe and fatal crashes, a number of useful insights are revealed. Figure 4 illustrates the items' (risk factors) relative frequencies in severe and fatal injuries, i.e. which categories of each independent variable are more common. Firstly, it is observed that clear weather conditions are the most frequent factor observed. However its statistical significance cannot be confirmed, since clear weather conditions are observed in the majority of SI and KSI crashes as well (see Table 2).

***** Please insert Figure 4 about here *****

***** Please insert Table 2 about here *****

The second most observed item (risk factor) in severe and fatal injuries, is the specific location of crashes and particular on segments different than intersections (No_intersection). The rest of the most frequently observed factors are "Male", "Urban", "Day", "PTW-Bicycle", "Night", "Car", "Non-Urban area" and "Runoff-Roll-fixed-other". Similarly, there are also differences in severity levels when the variables "Time" (night vs day) and "Area type" (urban vs non-urban) are considered. For instance, Table 3 shows that 885 KSI cases have occurred during the day, whilst 844 during the night. On the other hand, 7251 SI cases occurred during the day and 4451 during the night. Table 4 shows that although more KSI children/adolescents are observed in urban areas, the ratio of SI to KSI is much higher than non_urban areas. It is noted that the Eclat analysis should be interpreted with care, as it provides an illustration of the prevailing conditions of KSI cases but without any statistical justification and therefore it should be used as a supplementary analysis.

***** Please insert Table 3 about here *****

***** Please insert Table 4 about here *****

Tables 5 and 6 provide a visualization of the first 10 rules with the highest support (proportion) sorted (3 and 5 combinations of items-factors were used). The column "support" provides the proportion of observations with the respective combination of items (factors), while the column 'Count" is the frequency of each one.

***** Please insert Table 5 about here *****

***** Please insert Table 6 about here *****

Because crashes and injuries are a combination of more than a single factor, but rather a combination of multiple factors, the results of the Eclat model could provide good supplementary insights. To sum up, the most common combination of factors in KSI cases is the Clear weather-Male-No intersection-PTW/Bicycle-Urban area, followed by the Clear weather- Male-Night-No intersection-PTW/Bicycle, Clear weather-Male-Night-PTW/Bicycle-Urban area, Clear weather-Male-Night-No intersection-Urban area and Clear weather-Day-Male-No intersection-Urban area.

4.2 Mixed logit model

Afterwards, the mixed logit model was applied in order to analyse the impacts of crash and individual specific variables on injury severity of children and adolescents. 200 Halton draws were used, whilst the constant term was set as random following a normal distribution. The normal distribution was found to be the best, as this was the model with the best goodness of fit. In addition, a number of studies in the field (Behnood and Mannering, 2017; Moore et al., 2011) stated that the normal distribution is the most appropriate for the random parameters. Summary results of the model are presented on Table 7.

***** Please insert Table 7 about here *****

Overall, the mixed logit model showed good fit and was superior to the simple fixed effects model. More specifically, the log-likelihood ratio test of the null model versus the developed model was found to be statistically significant at a 95% level (p-value <0.001), as the log-likelihood value of the null model was -5157.1 and the respective value of the mixed logit model was -1620.3. On the other hand, the log-likelihood value of the fixed effects binary logistic model was -4727. Hence, the likelihood ratio between the fixed effects and the mixed logit model is significant at a 95% level, indicating that the mixed logit model outperforms the fixed effects model. An additional comparison using the Akaike Information Criterion (AIC) shows similar results (the lower AIC indicates a better fit). The AIC of the mixed logit model was found to be 3286.45, whilst the AIC of the fixed effects model was 9498. Consequently, the mixed logit model is preferred over the fixed effects model.

A first inspection on the specific model results, confirmed the hypothesis that there is unobserved heterogeneity among Greece counties on a spatial degree, because the standard deviation of the constant term is statistically significant at a 95% level (p-value <0.001). The temporal evolution is confirmed to some extend due to the fact that the variable “Year” has some statistically significant categories, indicating a downward trend especially in recent years. Firstly, the temporal instability in crash injury severities is something also indicated in the past (Behnood and Mannering, 2016, 2015). Moreover, there was a big effort in Greece during the last decade and a building of hundreds of kilometers of new motorways took place despite the economic crisis.

Regarding the person characteristics, it was found that males are usually associated with more severe injuries than females (beta coefficient = 0.261), something that can be considered to be in line with a few past studies in the field (Koopmans et al., 2015). However, Skjerven-Martinsen et al. (2014) did not find any significant influence of gender when analysing AIS 2+ injuries (moderate to severe injuries with at least one body section defined as injured) of child occupants. Another interesting remark of our study is that crashes involving pedestrians are associated with lower child/adolescent severity than head-on collisions and run-off-road collisions/collisions with fixed objects. This indicates that crashes involving teen drivers/ passengers tend to be more severe.

As for the crash environment characteristics, interesting trends were generally observed. Firstly, the positive sign of the beta coefficient shows that night crashes are usually more severe than day crashes. This finding is similar to Koopmans et al. (2015), who state that daylight was associated with less severe injuries of child pedestrians. Moreover, another study by Skjerven-Martinsen et al. (2014), showed that among multi-vehicle crashes occurring during daylight, 12% of child occupants sustained AIS 2+ injuries, whilst the respective percentage while traveling in darkness is 38%. It is noted that past research in Greece has shown the same impact of day and night on crash severity (Theofilatos et al., 2012). On the other hand, weather conditions were not found to have an influence on injury severity, a finding that is also observed in a few of similar past studies in Greece (Theofilatos, 2017). However, a few other past studies in Greece indicated no impact of weather (Theofilatos et al., 2012).

Interestingly, urban crashes are more likely to be less severe than crashes outside urban areas, similar to Peek-Asa et al. (2010), who suggested that fatality rates were generally higher for rural than urban teen drivers for all causes. This finding could be attributed to lighting conditions or increased driving speeds present on rural areas or highways. Additionally, crashes occurring at intersections are associated with slighter injury outcomes than crashes outside intersections. These findings could be considered to justify the previous result that children and adolescents are more vulnerable as drivers/passengers during trips outside urban areas than being pedestrian on urban areas. This result is considered interesting and needs further research.

Lastly, vehicle type was also found to have an influence on injury severity. As intuitively expected and also consistent with Theofilatos et al. (2012), child and adolescents involved in Two-Wheeler crashes (bicycles, mopeds and motorcycles) were more likely to sustain severe and fatal injuries than those involved in car and bus crashes. In our study, children and adolescents are more likely to be slightly injured in bus crashes compared to car crashes, in contrast to a study by García-España and Durbin (2008), which suggested that injury risk of car occupants is higher than SUVs, large vans and minivans (but not pickup trucks). The findings emphasize the need to raise awareness towards young people especially when they ride or are passengers of Two-Wheelers and cars.

Overall, although a number of findings may sound contradictory to the previous results of the AR model, they offer insights on the prevailing conditions when fatal and severe injuries of children and adolescents occur. However, when differences are observed it is encouraged to visually inspect the cross tabulation analysis of injury severity and each individual risk factor.

5. Conclusions

Despite the fact that a lot of studies have analysed the factors contributing to crashes including children and adolescents, relatively little research has been carried out regarding the injury severity analysis of children and adolescents as passengers and pedestrians. Our study uses a large national sample of disaggregate road crash data from all the regions in Greece for the period 2006-2015, including a total of 13431 total reported injuries. In order to account for unobserved heterogeneity both in a temporal and a spatial degree, a mixed logit model where the constant term followed a normal distribution is applied in which the year was also included as an independent variable. Moreover, to enrich discussion an Association Rule learning (AR) model was applied to explore the most frequent combinations of risk factors in severe and fatal injuries.

The estimation results showed that night crashes, crashes outside urban areas, run-off-road collisions/collisions with fixed objects as well as crashes involving males, bicycles or Powered-Two-Wheelers (PTWs) are associated with higher injury severity of children and adolescents. Interestingly, crashes involving pedestrians are associated with lower child/adolescent severity than head-on collisions and run-off-road collisions and collisions with fixed objects. This might indicate that that crashes where teen drivers are involved as drivers are likely to be more severe.

The aforementioned results show a need for focusing on policy measures such as law enforcement, especially when adolescents are riding a bicycle/PTW or when children/adolescents are involved during night trips outside urban areas. Awareness raising and campaigns as well as intervention programs to increase safe teen driving outside urban areas are also needed to address potential risk factors associated with rural and freeway environments. Additionally, although it is not explicitly obvious by the current study findings, there is a need to launch campaigns to ensure that child restraint is properly used, since such measures have been found to have positive effect on child restraint used (Aigner-Breuss and Pilgerstorfer, 2017).

The findings of the Association Rule (AR) learning provided additional insights on the combination of prevailing risk factors in fatal and severe injuries and serve as a supplementary data exploration technique; however results should be interpreted with care as in some cases contradictory findings are observed. Furthermore, the AR model has the disadvantage that it does not provide statistical significant estimates and probabilities. In some cases, a cross-tabulation analysis will further assist and it is something strongly suggested in such types of analysis. Therefore, the authors suggest to utilize both approaches.

Regarding further research, it would also be interesting to develop similar models for adults involved in road crashes in Greece and compare the results with the present paper's findings. Although the utilized dataset was rich, more in-depth data that would enhance understanding the manoeuvres, kinematics and driver behaviour of crashes involving young people are needed as well. Nevertheless, the findings of this study can assist in unveiling crash risk factors and prioritize programs and measures to promote road safety of children and adolescents in Greece, so as to tackle this problem even though there is a decreasing trend in the total number of injuries each year in Greece.

References

Adesunkanmi, A.R.K., Oginni, L.M., Oyelami, O.A., Badru, O.S., 2000. Road traffic accidents to African children: Assessment of severity using the Injury Severity Score (ISS). Injury, 31: 225–228.

Aigner-Breuss, E., Pilgerstorfer, M, 2017. Awareness raising and campaigns – Child restraint, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu on 17/09/2019.

Arbogast, K.B., Ghati, Y., Menon, R.A., Tylko, S., Tamborra, N., Morgan, R.M., Arbogast, K.B., Ghati, Y., Menon, R.A., Tylko, S., Tamborra, N., Morgan, R.M., Investigation, F., 2005. Field Investigation of Child Restraints in Side Impact Crashes Field Investigation of Child Restraints in Side. Traffic Injury Prevention, 6: 351–360.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research, 16: 35–47.

Behnood, A., Mannering, F.L., 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic Methods in Accident Research, 12: 1–17.

Behnood, A., Mannering, F.L., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. Analytic Methods in Accident Research, 8: 7–32.

Bennet, S.A., Yiannakoulias, N., 2015. Motor-vehicle collisions involving child pedestrians at intersection and mid-block locations. Accident Analysis and Prevention, 78: 94–103.

Brown, J., Hatfield, J., Du, W., Finch, C.F., Bilston, L.E., 2010. The Characteristics of Incorrect Restraint Use Among Children Traveling in Cars in New South The Characteristics of Incorrect Restraint Use Among Children Traveling in Cars in New South. Traffic Injury Prevention, 11: 391–398.

Brown, J.K., Jing, Y., Wang, S., Ehrlich, P.F., 2006. Patterns of severe injury in pediatric car crash victims: Crash Injury Research Engineering Network database. Jouranl of Pediatriatric Surgery, 41: 362–367.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research, 3–4: 11–27.

Das, S., Kong, X., Tsapakis, I., 2019. Hit and run crash analysis using association rules mining. Journal of Transportation Safety and Security, doi: 10.1080/19439962.2019.1611682.

Decina, L.E., Lococo, K.H., 2005. Child restraint system use and misuse in six states. Accident Analysis and Prevention, 37: 583–590.

Durbin, D.R., Chen, I., Smith, R., Elliott, M.R., Winston, F.K., Currently, A.B., 2005. Effects of Seating Position and Appropriate Restraint Use on the Risk of Injury to Children in Motor Vehicle Crashes. Pediatrics, 115: 305–309.

Duddu, V.R., Kukkapalli, V.M., Pulugurtha, S.S. 2019. Crash risk factors associated with injury severity of teen drivers. IATSS Research, 43: 37–43.

García-España, J.F., Durbin, D.R., 2008. Injuries to belted older children in motor vehicle crashes. Accident Analysis and Prevention, 40: 2024–2028.

Hagel, B.E., Romanow, N.T.R., Enns, N., Williamson, J., Rowe, B.H., 2015. Severe bicycling injury risk factors in children and adolescents: A case-control study. Accident Analysis

and Prevention, 78: 165–172.
Hahsler, M., Grun, B., Hornik, K., 2005. Association Rules and Frequent Itemsets. Journal of Statistical Software, 14: 1–6.
Kim, J., Ulfarsson, G.F., Shankar, V.N., Mannering, F.L., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accident Analysis and Prevention, 42:1751–1758.
Kim, M., Kho, S., Kim, D., 2017. Hierarchical ordered model for injury severity of pedestrian crashes in South Korea. Journal of Safety Research, 61: 33–40.
Koekemoer, K., Van Gesselleen, M., Van Niekerk, A., Govender, R., Van As, A.B., 2017. Child pedestrian safety knowledge, behaviour and road injury in Cape Town, South Africa. Accident Analysis and Prevention, 99: 202–209.
Koopmans, J.M., Friedman, L., Kwon, S., Sheehan, K., 2015. Urban crash-related child pedestrian injury incidence and characteristics associated with injury severity. Accident Analysis and Prevention, 77:127–136.
Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11: 1-16.
Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accident Analysis and Prevention, 43:621–630.
Olsen, C.S., Cook, L.J., Keenan, H.T., Olson, L.M., 2010. Driver seat belt use indicates decreased risk for child passengers in a motor vehicle crash. Accident Analysis and Prevention, 42: 771–777.
Peek-Asa, C., Britton, C., Young, T., Pawlovich, M., Falb, S., 2010. Teenage driver crash incidence and factors influencing crash injury by rurality. Journal of Safety Research, 41: 487–492.
Pour-Rouholamin, M., Zhou, H., 2016. Investigating the risk factors associated with pedestrian injury severity in Illinois. Journal of Safety Research, 57: 9–17.
Shaheed, M.S., Gkritza, K., 2014. Analytic Methods in Accident Research A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research, 2: 30–38.
Skjerven-Martinsen, M., Naess, P.A., Hansen, T.B., Gaarder, C., Lereim, I., Stray-Pedersen, A., 2014. A prospective study of children aged <16 years in motor vehicle collisions in Norway: Severe injuries are observed predominantly in older children and are associated with restraint misuse. Accident Analysis and Prevention, 73:151–162.
Theofilatos, A., 2017. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. Journal of Safety Research, 61: 9–21.
Theofilatos, A., Graham, D., Yannis, G., 2012. Factors Affecting Accident Severity Inside and Outside Urban Areas in Greece. Traffic Injury Prevention, 13: 458–467.
Theofilatos, A., Ziakopoulos, A., 2018. Examining Injury Severity of Moped and Motorcycle Occupants with Real-Time Traffic and Weather Data. Journal of Transportation Engineering Part A Systems, 144: 1–12.
Viano, D.C., Parenteau, C.S., Viano, D.C., Parenteau, C.S., 2008. Fatalities of Children 0 – 7 Years Old in the Second Row Fatalities of Children 0 – 7 Years Old in the Second Row. Traffic Injury Prevention, 9, 231–237.
Washington, S.P., Karlaftis, M.G., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis. CRC Press.
Who Unicef, 2008. Children and road traffic injury. Injury Prevention, 2.
Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. Accident Analysis and Prevention, 121: 285–294.

Xu, X., Xie, S., Wong, S.C., Xu, P., Huang, H., Pei, X., 2016. Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model. Journal of Advanced Transportation, 50, 2015–2028.

Zaveri, P.P., Morris, D.M., Freishtat, R.J., Brown, K., 2009. Overweight children: Are they at increased risk for severe injury in motor vehicle collisions? Accident Analysis and Prevention, 41, 959–962.

Zhai, X., Huang, H., Sze, N.N., Song, Z., Hon, K.K., 2019. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. Accident Analysis and Prevention, 122: 318–324.

List of tables and figures

Table 1. Variables considered, potential values and descriptive statistics

Variable	Values	Number of cases
Severity	Slighly injured	11702
	Killed/Severely injured	1729
Gender	Female	4794
	Male	8637
Illumination	Day	8136
	Night/Dusk	5295
Crash Type (collision type)	Head-on	1016
	With pedestrian	3650
	Rear-end	810
	Off road/Fixed object/Other	2423
	Side	4837
	Sideswipe	695
Weather	Adverse	1038
	Clear	12393
Area type	Urban	10492
	Non-urban	2939
Intersection	Yes	5177
	No	8254
Vehicle Type involved	Bus/Truck	643
	Car	7111
	PTW-Bicycle	5478
	Other	199

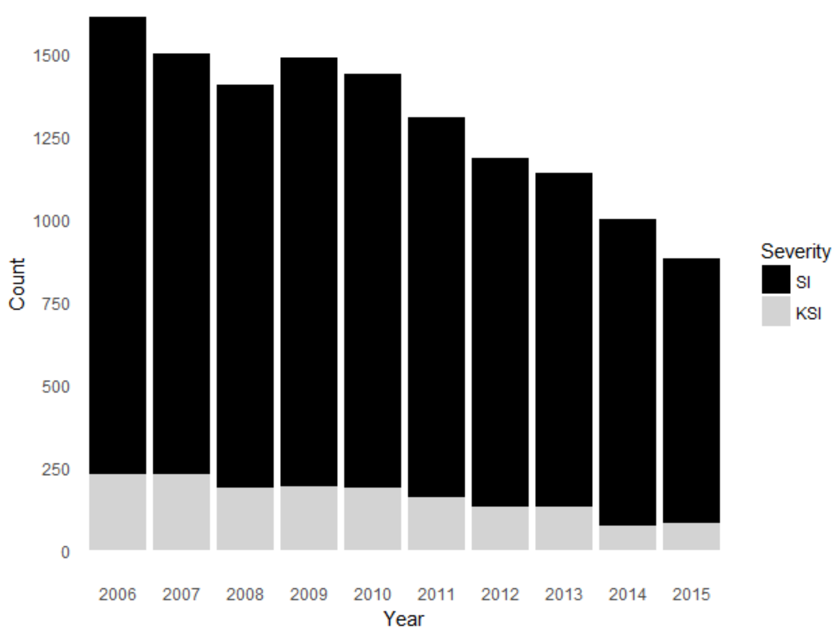

Figure 1. Temporal evolution of SI and KSI child and adolescent cases in Greece (2006-2015)

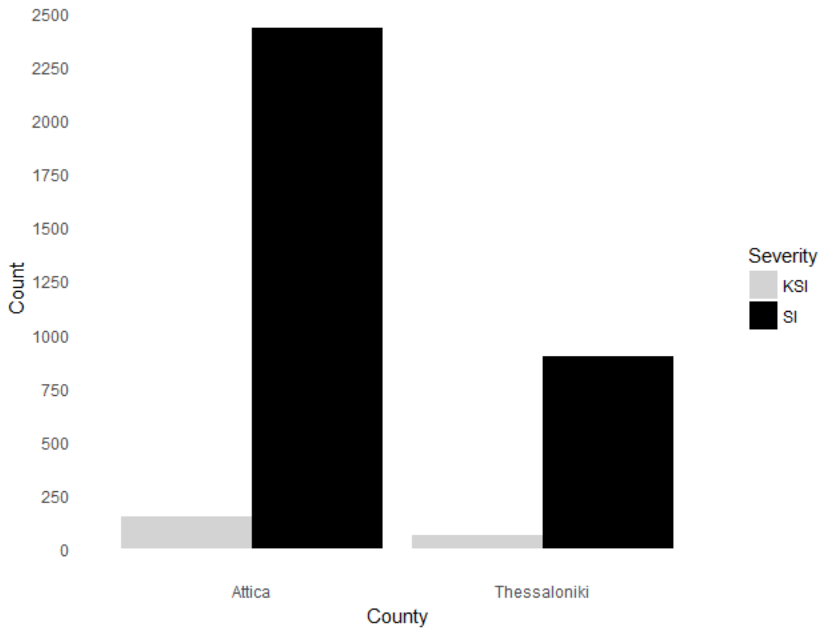

Figure 2. Distribution of SI and KSI of child and adolescent cases in the counties of Attica and Thessaloniki of Greece (2006-2015)

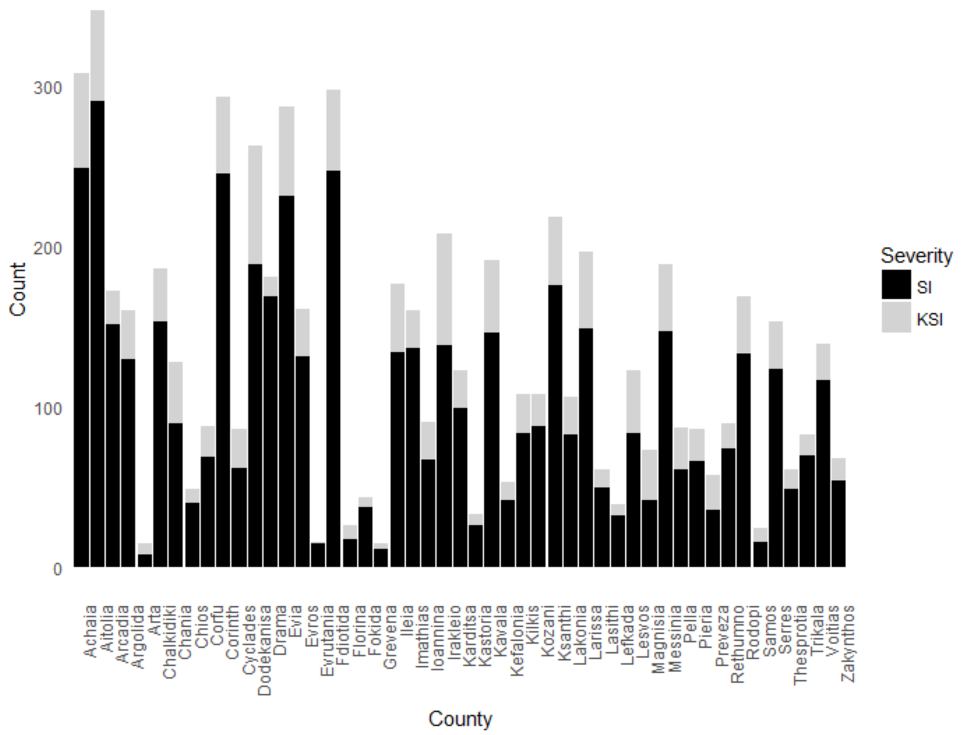

Figure 3. Distribution of SI and KSI of child and adolescent cases in the rest of Greek counties (2006-2015)

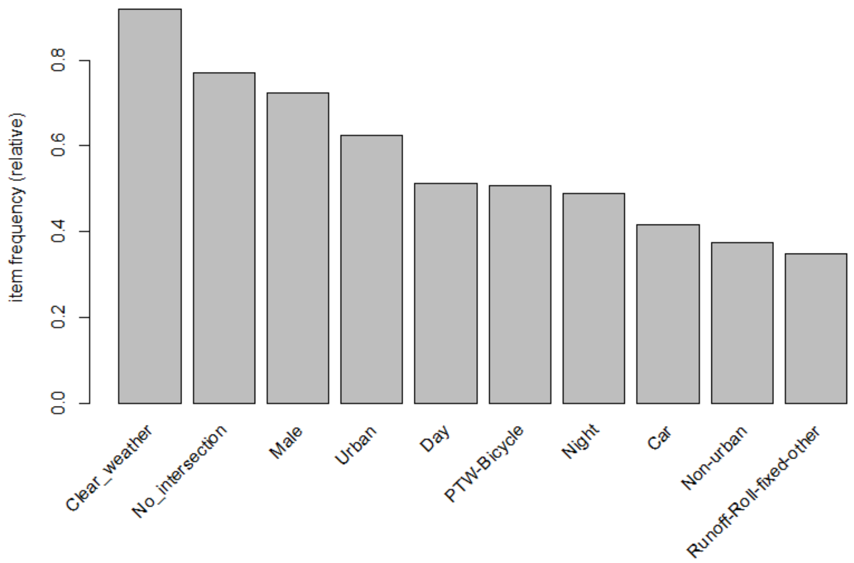

Figure 4. Item (risk factors) relative frequencies in KSI cases.

Table 2. Frequency distribution of weather conditions for SI and KSI cases

Severity	Adverse_Weather	Clear_weather
SI	898	10804
KSI	140	1589

Table 3. Frequency distribution of time and for SI and KSI cases

Severity	Day	Night
SI	7251	4451
KSI	885	844

Table 4. Frequency distribution of location for SI and KSI cases

Severity	Non_urban	Urban
SI	2290	9412
KSI	649	1080

Table 5. Combinations of 3 of the most prevailing factors in KSI cases -the first 10 rules with the highest supports (sorted)

Combination of 3			
A/A	Items	Support (proportion)	Count
1	{Clear weather, Male, No intersection}	0.5087	880
2	{Clear weather, Male, Urban area}	0.4277	740
3	{Clear weather, No intersection, Urban area}	0.4087	707
4	{Clear weather, Male, PTW-Bicycle}	0.4081	706
5	{Clear weather, Day, No intersection}	0.3595	622
6	{Clear weather, PTW-Bicycle, Urban area}	0.3462	599
7	{Clear weather, Night, No intersection}	0.3439	595
8	{Clear weather, Day, Male}	0.3376	584
9	{Clear weather, No intersection, PTW-Bicycle}	0.3353	580

10	{Clear weather, Male, Night}	0.3295	570

Table 6. Combinations of 5 of the most prevailing factors in KSI cases-the first 10 rules with the highest supports (sorted)

	Combination of 5		
A/A	Items	Support (proportion)	Count
1	{Clear weather, Male, No intersection, PTW-Bicycle, Urban area}	0.1815	314
2	{Clear weather, Male, Night, No intersection, PTW-Bicycle}	0.1636	283
3	{Clear weather, Male, Night, PTW-Bicycle, Urban area}	0.1566	271
4	{Clear weather, Male, Night, No intersection, Urban area}	0.1543	267
5	{Clear weather, Day, Male, No intersection, Urban area}	0.1416	245
6	{Clear weather, Day, Male, PTW-Bicycle, Urban area}	0.1295	224
7	{Clear weather, Male, Night, No intersection, Runoff-Roll-fixed-other}	0.1283	222
8	{Clear weather, Night, No intersection, PTW-Bicycle, Urban area}	0.1277	221
9	{Clear weather, Male, No intersection, PTW-Bicycle, Runoff-Roll-fixed-other}	0.1231	213
10	{Clear weather, Day, Male, No intersection, PTW-Bicycle}	0.1225	212

Table 7. Results of the mixed logit model

Variable	Beta coefficient	Standard error	p-value	Significance
Constant	-1.060	0.163	<0.001	at 99%
Sd of constant	0.100	0.028	<0.001	at 99%
Gender_female (reference category)	-	-	-	-
Gender_male	0.261	0.068	<0.001	at 99%
Crash_type-Head_on (reference category)	-	-	-	-
Crash_type_Pedestrian	-0.339	0.105	<0.001	at 99%
Crash_type_Rear_end	-0.688	0.146	<0.001	at 99%
Crash_type_Run_off_road_fixed_object	0.307	0.096	0.001	at 99%
Crash_type_Side	-0.495	0.100	<0.001	at 99%
Crash_type_Sideswipe	-0.699	0.158	<0.001	at 99%
Time_Day (reference category)	-	-	-	-
Time_Night/Dusk	0.299	0.054	<0.001	at 99%
Area_type_Non-urban (reference category)	-	-	-	-
Area_type_Urban	-0.767	0.068	<0.001	at 99%
Intersection_No (reference category)	-	-	-	-
Intersection_Yes	-0.528	0.068	<0.001	at 99%
Vehicle_type_Bus/Truck (reference category)	-	-	-	-
Vehicle_type_Other/Unknown	0.209	0.231	0.366	non-significant
Vehicle_type_Car	-0.44	0.122	<0.001	at 99%
Vehicle_type_PTW_Bicylce	0.286	0.125	0.021**	at 95%
Year_2006 (reference category)	-	-	-	-
Year_2007	0.077	0.102	0.445	non-significant

Year_2008	-0.128	0.107	0.233	non-significant
Year_2009	-0.067	0.106	0.522	non-significant
Year_2010	-0.081	0.107	0.445	non-significant
Year_2011	-0.102	0.111	0.356	non-significant
Year_2012	-0.207	0.118	0.078	at 90%
Year_2013	-0.141	0.117	0.228	non-significant
Year_2014	-0.623	0.141	<0.001***	at 99%
Year_2015	-0.388	0.136	<0.001***	at 99%