A preliminary analysis of in-depth accident data for powered two-wheelers and bicycles in Europe

Ziakopoulos A., Theofilatos A., Yannis G., Margaritis D., Thomas P., Morris A.,
Brown L., Robibaro M., Usami, D. S., Phan, V., Davidse R., Buttler, I.

Abstract Despite progress from scientific and technological advancements, road safety remains a major issue worldwide. Road accident impacts such as fatalities, injuries and property damage consist considerable costs borne not only by involved people but society as well. This study aims to present preliminary findings of in-depth accident analysis for two-wheelers (bicycles and powered two wheelers – PTWs) across six countries in Europe. Data regarding the conditions underlying accident occurrence are presented, including time and date, weather, vehicle and road conditions and rider-related parameters such as age, intoxication and use of protective equipment. In addition, a Two Step Cluster Analysis is implemented in order to explore any possible classification of the analysed cases. It appears that two clusters are formed: the first includes more favourable conditions ("no wind, no drugs, good lighting") while the second consists of less favourable conditions for road safety ("windy, lighting, unknown DUI condition"). This hints at a meaningful separation of the examination of two-wheeler accidents when the influence of outside factors is considerable. The inclusion of different but representative areas across Europe offers robustness and transferability to the data and respective results.

Keywords: accident analysis, bicycle, in-depth investigation, ptw, random forest analysis, road safety.

I. INTRODUCTION

Despite considerable efforts and relevant progress, road safety remains a global issue worldwide. In 2016, 25,670 people lost their lives on EU roads [1], while in recent years the number of road traffic deaths globally has plateaued at 1.25 million a year [2]. Identifying the causes of accident occurrence is a necessary step in order to create new effective road safety measures and policies, and break through the stalemate, which has been one of the main fields of interest of road safety researchers worldwide [3, 4]. Thus the objective of the present study is the investigation and ranking of causes of accident occurrence by conducting a Random Forest Analysis and to group accidents involving powered two wheelers and bicycles into meaningful groups (clusters) after a descriptive statistic evaluation and overview of the collected in-depth microscopic accident data.

A number of studies have been published in the literature for two-wheeler safety, especially for PTWs, regarding the correlation of injury severity with a number of external variables such as weather conditions, road geometry, speeding, alcohol and others. Results mainly show that a large number of these variables do influence road accident severity for PTWs considerably. Examples include negative influence for accidents in junctions [5], while speeding [6] and on pavement surface [7], while in darkness [8], and for specific accident types [9-10]. There are also overcompensation effects for adverse weather conditions resulting in more conservative driving [7, 11]. Vehicle age and lack of helmet use have been found to play an impact on increased accident severity as well [12]. Another proposed important factor appears to be PTW interaction with car drivers [13]. More recently, another study found 1-hour traffic flow and variations in speed have a significant influence on PTW involvement in accidents on urban arterials [14]. In many relevant studies, the variations of accident characteristics and the particularities of each incident are evident, and warrant a microscopic, in-depth investigation, namely further and more detailed examination of two-wheeler accident data is warranted.

Apostolos Ziakopoulos is a Research Associate, Athanasios Theofilatos is a Postgraduate Researcher and George Yannis is a Professor in the Department of Transportation Planning and Engineering of the National Technical University of Athens. Dimitrios Margaritis is a Research Associate in the Centre for Research & Technology of the Hellenic Institute of Transport. Pete Thomas is a Professor, Andrew Morris is a Professor and Laurie Brown is a Research Associate in the Transport Safety Research Group of Loughborough Design School. Massimo Robibaro and Davide Shingo Usami are Research Associates in the Centre for Transport and Logistics at the University of Rome "La Sapienza". Vuthy Phan is a Research Associate in the Centre European Studies Safety And Analysis Des Risques. Ragnhild Davidse is a Research Associate in SWOV Institute for Road Safety Research. Ilona Buttler is a Research Associate in Instytut Transportu Samochodowego.

II. METHODS

A. Data collection and processing

The aim of the study is achieved by analysing in-depth microscopic data from numerous two-wheeler accidents from six sample regions across Europe (France, Greece, Italy, The Netherlands, Poland, and The United Kingdom). Several databases have been constructed regarding the vehicle, road environment and the road users themselves, each including numerous detailed variables.

For this study, the methodology protocol established in DaCoTa was followed [15, 16] (in terms of data structure and data collection procedures), which was developed by the partners of that endeavor. This protocol entails a large number (over 100) of parameters that convey important information for the road environment, involved vehicles and persons, circumstances and causes for each accident etc.

The DaCoTa sampling methodology states that the sample should "closely represent all types of traffic accidents occurring on the public roads, adequately covering all hours of the day, all days of the week and all levels of injury severity." Therefore, based on the protocol, a main purpose of this study was to select accidents by paying attention at diversifying the accident conditions, in particular crash location, severity and time period. Additionally, considerable effort was made for the DaCoTa protocol to be modified for the needs of the study in order to focus on two-wheelers. The parameter information platform was adjusted to include more information for specialized two-wheeler aspects as well.

To that end, in-depth motorcycle, moped and bicycle accident data was collected from discrete sampling areas within the six European countries. The selected sample regions covered by each team had a known relationship with the respective national accident populations. The distribution of key variables was intended in principle to be close to that of the national data, for the data to be representative. An example is shown for the sampling area in the Netherlands on Table 1. Similar criteria were used for bicycles, and in other countries as well.

Table 1: Road accidents with at least a PTW involved (The Netherlands, 2010-2012)

Criterion	% in data collection area*	% in country*
Fatal accident	2%	4%
Injury accident (non-fatal)	98%	96%
Truck involved	2%	2%
Pedestrian involved	3%	4%
Motorway	2%	4%
Rural	14%	18%
Urban (but not on Motorway)	67%	65%
Unknown	16%	14%
January – April	33%	34%
May – August	45%	42%
September – December	22%	24%
08:00 – 19:59	78%	76%
20:00 – 07:59	22%	23%
Total number of injury accidents (100%)	969 (3 years)	6.585 (3 years)

* Percentage calculated with reference to total injury accidents.

The in-depth accident investigations were categorized into two main groups: investigations conducted at the scene of the accidents (within minutes to hours of accident occurrence) or investigations conducted retrospectively (within a few days of the accident occurrence). The purposive sampling was based on the concept of saturation, defined as the point at which the data collection process no longer offers any new or relevant data. In total, 500 cases comprising 77% PTW and 23% bicycle accidents occurring in 2015 and 2016 were investigated.

For the initial analysis of the data, descriptive statistics have been used. Parameters were cross-tabulated in multiple layers when needed, in order to provide the proportional differences and to underline the dominant factors that appear most frequently in accident scenarios.

The graphical presentation of the findings contains tables, pies and bar graphs, as well as pictograms that illustrate more clearly the accident configurations and the contributing accident factors.

B. Two-step clustering

In order to further investigate the data towards the aim of the study, a Two Step Cluster Analysis was carried out on the dataset where one case represented a recorded accident. This method of clustering is able to produce solutions based on both continuous and categorical variables. The clustering algorithm is based on a distance measure. The first step of the two-step procedure is the formation of pre-clusters. The goal of pre-clustering is to reduce the size of the matrix that contains distances between all possible pairs of cases. In the second step, the standard hierarchical clustering algorithm is applied on the pre-clusters.

Regarding the distances between clusters, the log-likelihood distance measure is selected as appropriate for both continuous and categorical variables, (as opposed to the Euclidean distance handles only continuous variables). This distance is probability based, and expresses the decrease in log-likelihood as sub-clusters are combined into one bigger cluster. For the two-step algorithm, the distance is dependent on the total numbers of continuous and categorical variables, their respective variances and the total number of data records. This method assumes normal distributions for continuous variables when calculating log-likelihoods (in practice, this means that the two-step cluster algorithm requires that all continuous variables are standardised). The clustering criterion (in this case the BIC – Bayesian Information Criterion) is computed for each potential number of clusters. Smaller values of the BIC indicate better clustering outcome. Also, a satisfactory solution should have a large ratio of BIC Changes and a large ratio of distance measures.

C. Random Forest Analysis

As an in-depth follow-up to the initial statistical analysis, insight was sought on the parameters affecting injury severity of occupants (dead against no-dead). For that reason, the Random Forest analysis was chosen in order to rank the explanatory variables according to their relative importance. A random forest is a classifier including a collection of tree-structured classifiers {h(x, Θk), k = 1,...}, where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x [17]. A relevant study [18] suggests a number of bootstrap samples from the original sample have to be drawn and afterwards a classification tree to each bootstrap sample has to be fitted (number of trees). In the current analyses 100 trees were used.

Random Forest Analyses have been previously used in traffic safety studies [19-22]. These studies utilized this method to rank the variable importance and thus select the appropriate variables, sometimes as a preparatory step before applying additional statistical models. In this study the Random Forest Analysis is utilized to determine variable importance rankings when examining injury severity of accidents. This is achieved by separating fatal (i.e. resulting in one or more deaths for involved individuals) and non-fatal accidents, and comparing the two groups.

III. RESULTS

Descriptive Statistics

The overview analysis is based on 500 collected accidents. More than 500 two-wheelers have been examined within the study overall since it was possible that a case involved multiple PTWs and/or an accident between a PTW and a bicycle. It is noteworthy that for the following analyses, the total number of cases for every graph might fluctuate by a small degree, since cases lacking the examined variable were excluded from the graphs. Results for accident severity distributions of cases among examined countries are shown on Table 2.

From Figure 1 it is evident that all different injury severity categories can be found in considerable percentages amongst accidents. This observation holds for both examined vehicle categories (whether two-wheelers are powered or not). It can be observed that across vehicles, serious injuries are more common than slight injuries, which are in turn more common than fatalities. This might hint that more severe accidents are more commonly collected for two-wheelers (especially in countries using retrospective methods). Injury severities for some cases were not recorded, while there have been some small additional numbers of two-wheeler riders which were not injured in the respective accident.

Table 2: Two-wheeler accident severity of investigated accidents

Country	Not Injured	Slight	Serious	Fatal	Unknown	Total
France	2	43	32	7	3	86
Greece	-	59	11	15	-	85
Italy	16	-	-	2	57	75
The Netherlands	2	-	84	1	-	87
Poland	-	59	19	7	2	87
United Kingdom	-	16	10	54	-	80
Total	**20**	**177**	**156**	**86**	**62**	**500**

Figure 2 shows that most PTW and bicycle accidents occur mainly in the period from May to September. This is something that stands to reason considering the relatively good weather in that period of time during which two-wheelers are used, and perhaps increased exposure due to tourism, especially in countries such as Greece and Italy. Similarly, as shown in Figure 3, regarding the day of the week, it can be observed that the majority of accidents (total, PTW, bicycles) occur in working days (Monday-Friday) which can be interpreted due to the higher number of trips and traffic flow observed in road networks. It should be mentioned at this point that the collected data do not include exposure information, and results should be treated likewise.

From Figure 4, it can be inferred that the proportion of accidents that occur within the day seems to follow an approximately normal distribution, which indicates that between 20:00 and 3:00 the least amount of accidents occur.

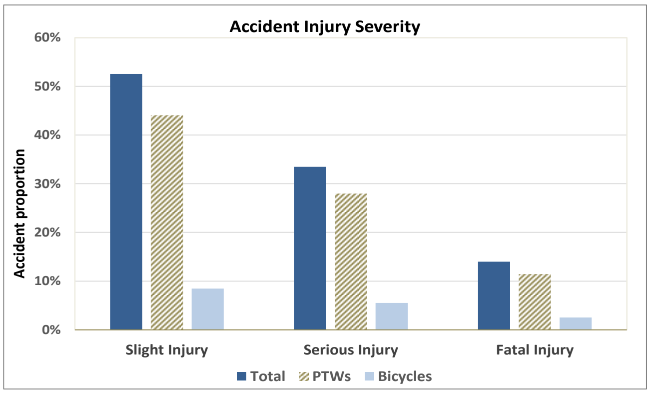

Fig. 1. Accident injuries by vehicle type

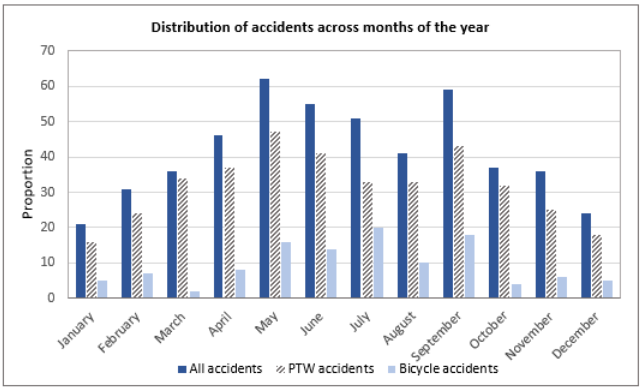

Fig. 2. Accidents throughout the year by vehicle type

Below follow analyses and statistics based on variables concerning human factors of road users. The database currently contains 951 road users, of which 780 are drivers or riders of the vehicle. 55 road users are passengers in vehicles, and 10 are pedestrians, while further updates and new case data are expected as well. Further characteristics than those presented are also available; however for 106 road users they were particularly complex to decode and are still being processed in order to fully complete the database.

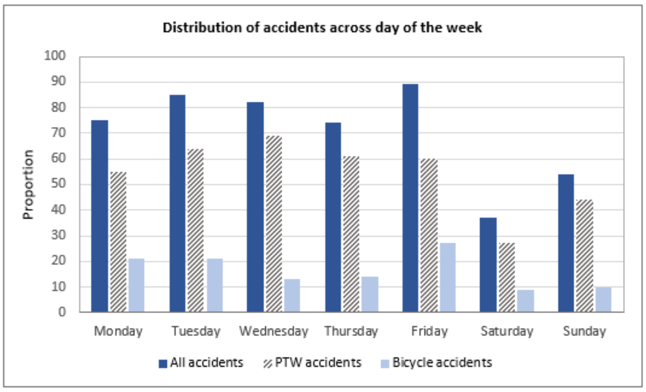

Fig. 3. Accidents throughout the week by vehicle type

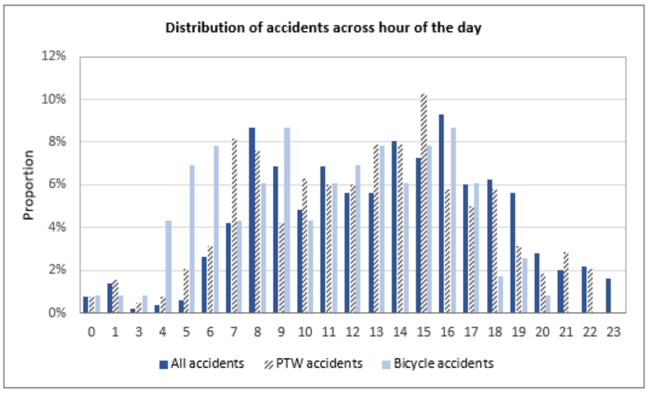

Fig. 4. Accidents throughout the day by vehicle type

It was found that the vast majority of accidents (>80%) have occurred without any alcohol involvement of drivers. Regarding medical history, it appears that most riders had not any recorded pre-existing medical condition.

Speeding is one of the most critical contributions for accident occurrence and severity [23, 24]. Figures 5 and 6 depict speeding contributions for PTW and other riders by injury and age respectively. Firstly, for the majority of cases, speeding was not recorded as the primary accident contributor. When it was so, PTW riders were above the respective speed limit. On the contrary, bicyclists were below the respective speed limit; however, the nature of bicycles can be said to involve their occupants in accidents when speed is less than the speed limit but relatively high. Younger riders appear to speed more frequently, which makes sense given their overall more aggressive attitudes. Speed is also directly correlated with increased injury severities (leading to fatal/serious injury accidents over slight/no injury accidents).

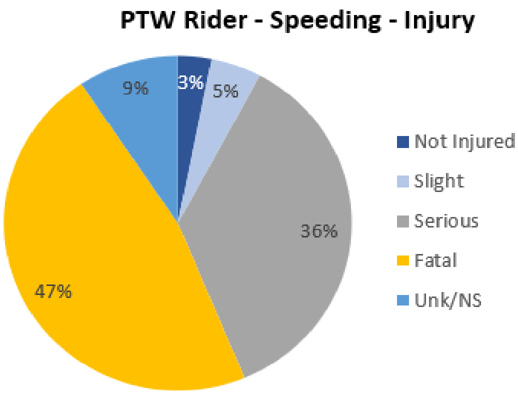

Fig. 5. Speeding distribution by injury

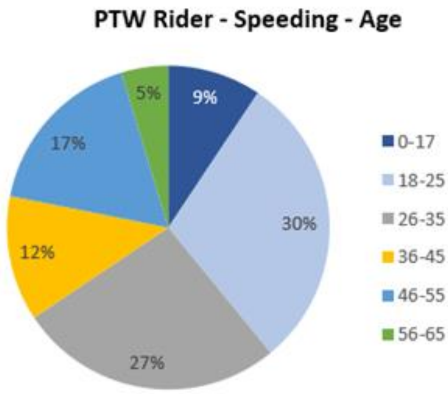

Fig. 6. Speeding distribution by age

Speeding parameters were examined further and are presented on Table 3, including findings for Other Interacting Road Users (OIRUs): (It should be noted that cases with missing data were excluded from the table)

Table 3: Speeding contribution for accidents

Speed is a contributing factor	PTW Riders	Percentage	Bicycle Riders	Percentage	OIRUs	Percentage
No	143	48%	85	77%	265	81%
Yes, and above speed limit	54	18%	2	2%	7	2%
Yes, and below speed limit	3	1%	4	4%	2	1%
Yes, and speed limit unknown	7	2%	0	0%	2	1%
Unknown	94	31%	20	18%	50	15%

Protective equipment such as helmets was examined specifically for PTW riders. While most riders seem to recognise the value of using helmets (81%), a non-negligible percentage did not (15%), hinting at margins for awareness raising for safety equipment. When excluding unknown cases, helmets did stay on during the majority of accidents, proving their high effectiveness levels. Results for cyclists showed that 36% of riders were wearing a helmet and had it fastened, and 39% were not wearing one at all. For 17% of the cases, a helmet was worn but not fastened, therefore indicating a misuse with not guaranteed rider protection.

When examining the recorded data, it would appear that most two-wheeler riders do not use reflective clothing at all times, as shown in Figure 7. This percentage is somewhat decreased for bicycle riders (from about 65% to below 50% of the respective totals). There is thus some room for reaching higher usage levels of reflective clothing from riders, though this needs to be examined with daylight exposure in mind as well.

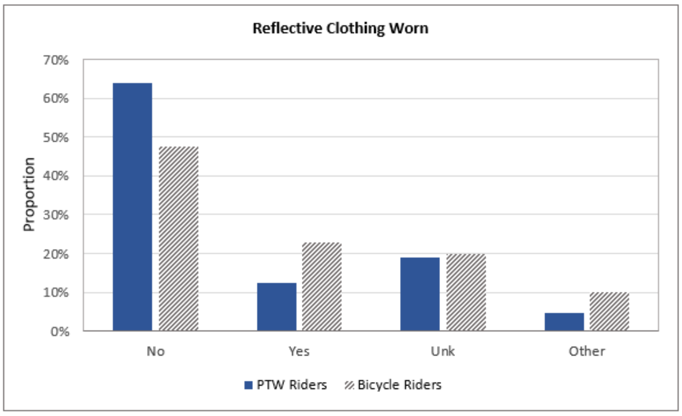

Fig. 7. Reflective clothing worn by two-wheeler riders

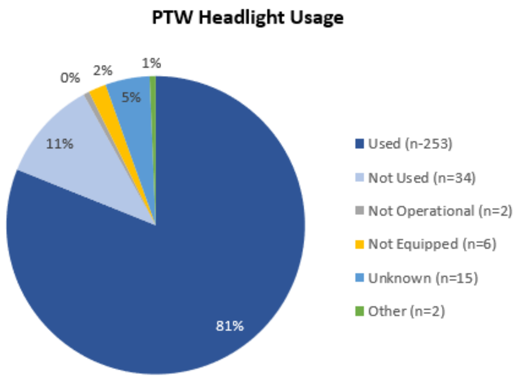

Fig. 8. PTW headlight usage

On the contrary, in the majority of cases (81%), PTWs riders did use their headlights, as shown in Figure 8. This is an observation across all examined countries, which have at times different visibility in their road environments. This can be affected by many factors (infrastructure elements, sight distance due to geometrical factors or other obstacles, darkness, weather effects) and by corresponding measures, for instance automated headlight or daylight running measures and regulations. Only 21% of cyclists had lights fitted and in use, suggesting that regulations such as mandatory daylight running lights that may result in higher PTW light usage.

Below follow analyses and statistics based on variables concerning the vehicle factors of vehicles involved in the examined accidents. The reasoning behind this analysis is to monitor the rates at which different elements of the PTWs deteriorate when examined in contrast to the overall condition of the vehicle (graded every time from excellent to poor, and shown in bars). As shown in Figures 9 and 10, in the majority of PTW cases, right and left mirrors are operational. In addition, there is considerable variation in the PTW condition when the brake lever is operational, meaning that mirrors are usually operational under all PTW conditions.

Figure 11 shows that when the front sprocket condition is in good condition then the PTW condition is also in excellent, good or at least fair condition. It is also shown that the poor PTW condition is mainly associated with worn sprocket, therefore this mechanical component seems related to overall vehicle condition. When the rear sprocket condition is in good condition then the PTW condition is also in excellent, good or at least fair condition, as shown in Figure 12. An interesting finding from Figure 13 is that when the throttle condition is in correct form, the condition of the vehicle varied considerably as high percentages of both excellent and poor PTW conditions are observed (more than 80%). When the steering stem adjustment is in correct form, the PTW condition is mainly in good condition but fair and poor conditions are also observed (about 50%), as shown in Figure 14.

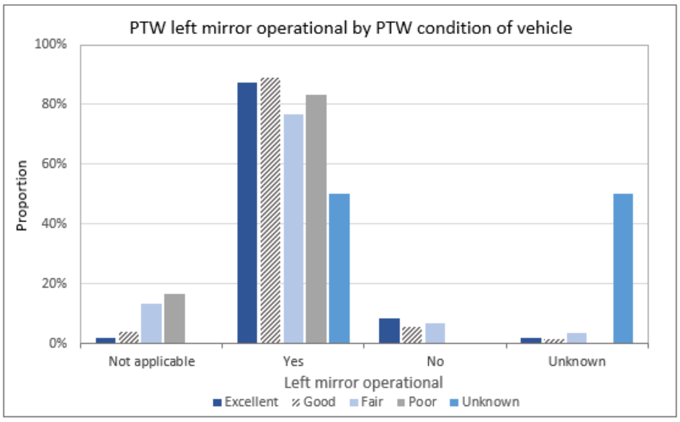

Fig. 9. PTW left mirror condition

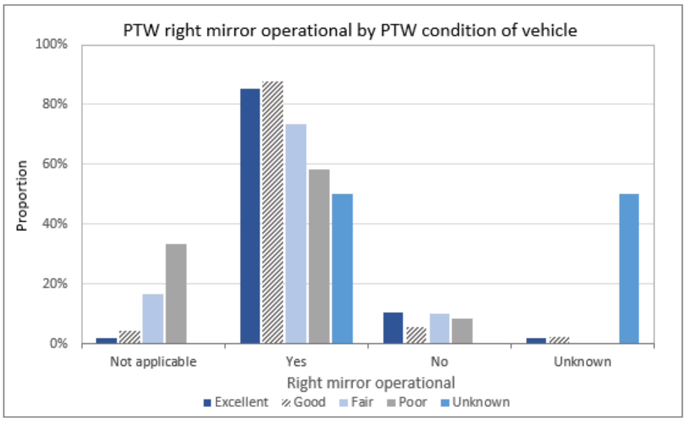

Fig. 10. PTW right mirror condition

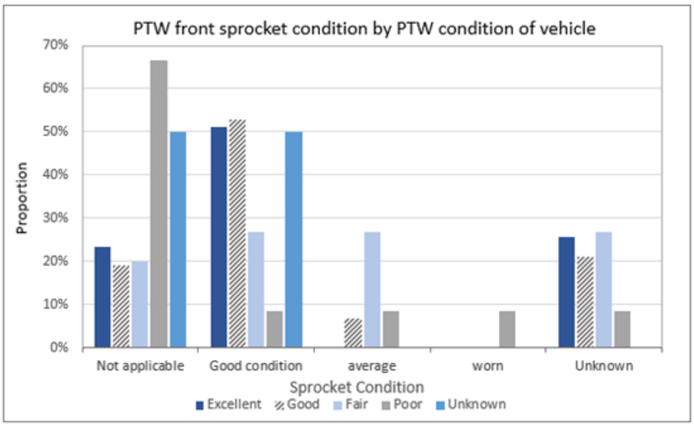

Fig. 11. PTW front sprocket condition

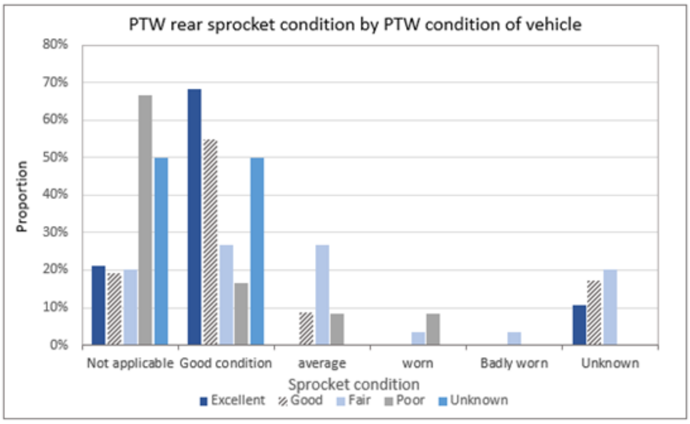

Fig. 12. PTW rear sprocket condition

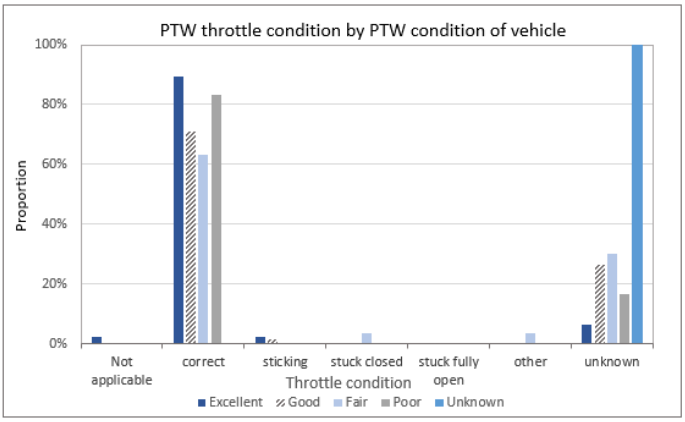

Fig. 13. PTW throttle condition

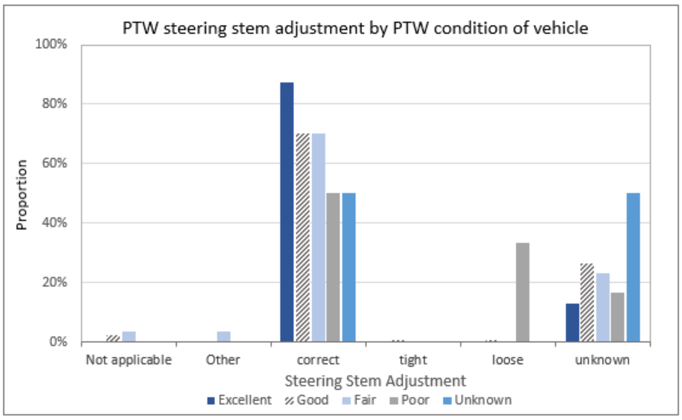

Fig. 14. PTW steering stem adjustment

The land use of the accident area was also examined. The predominant characteristics of the type of zone of the surrounding area were recorded for each of the 715 roads. It was found that the highest proportions of accidents occur in residential and commercial areas. More specifically, about 50% of PTW accidents occur in residential areas while 40% of bicycle accidents occur in commercial areas. Analysis results indicated that for all accident types, the majority occurred in urban areas as opposed to rural areas, and moreover PTWs had a higher proportion of rural accidents than bicyclists. It was found that regardless of the type of the vehicle (total, PTW, bicycle), about of 80% of accidents happen in daylight. The second most frequent light condition at the time of the accident is the electric light (about 11%).

In the following section, data concerning the road environment at the time of accident are providing, depicting overall accident conditions. The analysis was carried out on 715 roads across 500 recorded cases. For this section, the analysis is conducted on a mixture of “accident” and “road” level.

Data relevant to weather conditions are shown in Figures 15 and 16. In absolute percentages, the vast majority of accidents occur under good weather conditions regardless of the vehicle type (PTW, bicycle). Figure 16 below shows that fog was not present in almost all accident cases. On the other hand, although the majority of accident occurred under no wind, a non-negligible proportion occurred under 10-15km/h wind speed (it more than 10%) and under 15-20km/h (about 5%).

The majority of accidents occur in local roads, namely more than 35% of total and PTW accidents, as shown in Figure 17. It is interesting to see that a large share of bicycle accidents occur in collector roads (more than 30%) The fewer accidents occur in principal arterials. When conducting the analysis, for accidents occurring at junctions, the road with priority was designated as first road (priority defined by traffic signs, traffic lights, policemen or any other type of junction control). Almost half (50%) of any type of accident happens on single roads, as shown in Figure 18. More than about 20% of accidents happen on T or Y junctions whilst about 20% happen on crossroads. Accidents at roundabouts are less than 10%.

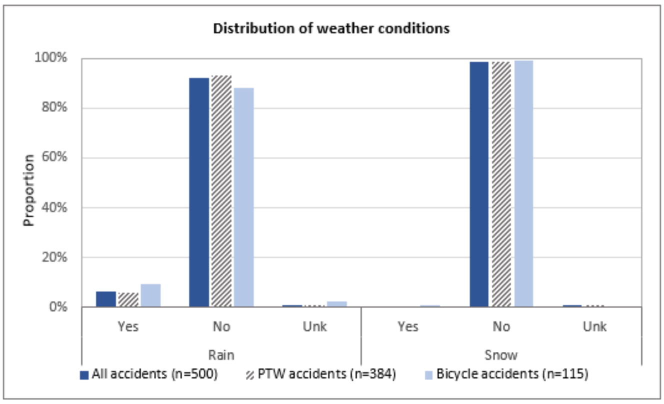

Fig. 15. Presence of rain and snow conditions

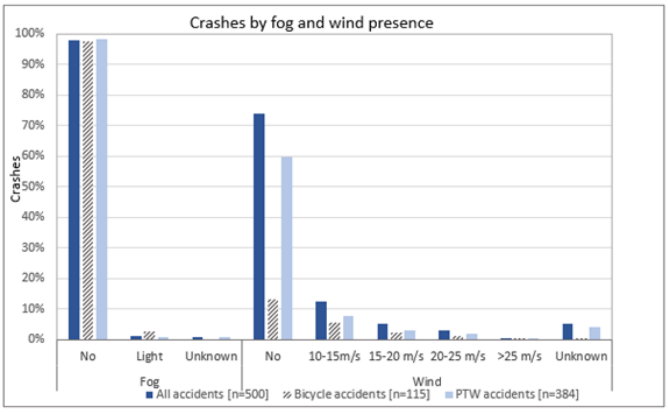

Fig. 16. Presence of fog and wind conditions

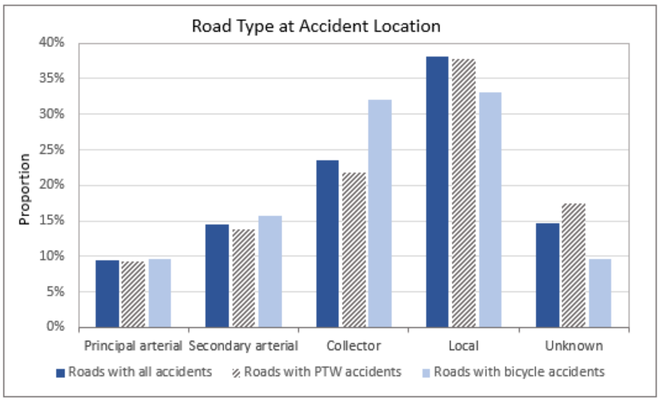

Fig. 17. Road types at accident locations

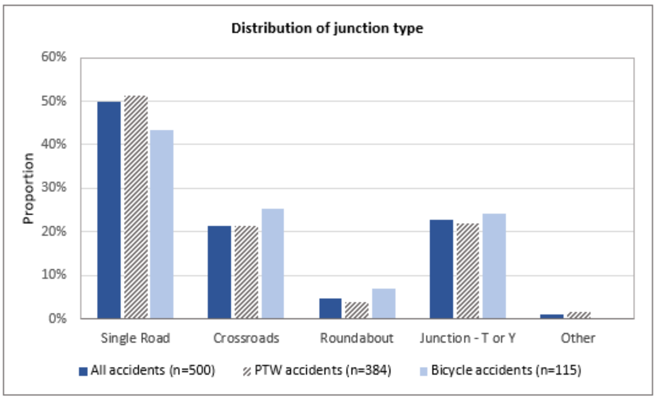

Fig. 18. Junction types at accident locations

Additionally, e-bikes were a rare vehicle category, which was treated separately for the analysis. They were found primarily in the Netherlands (11 of the 14 reported e-bikes in accidents). During the period between November and December a disproportionately high amount of accidents involving e-bikes was found.

Two Step Cluster analysis

This section presents the findings of the Two Step Cluster Analysis. As previously stated, it can be observed that two clusters were produced. The first cluster involves 398 accidents (cases) while the second involves 99 accidents (cases), as shown on Table 4. The overall model quality is considered good, as shown on Figure 19 which includes results from SPSS software, followed by an analysis of cluster profiles:

Table 4: Two Step Cluster Distributions

Characteristics		**N**	**Percentage of Combined**	**Percentage of Total**
Cluster	1	398	80.1%	79.6%
	2	99	19.9%	19.8%
	Combined	497	100.0%	99.4%
Excluded cases		3	-	0.6%
Total		**500**	**-**	100.0%

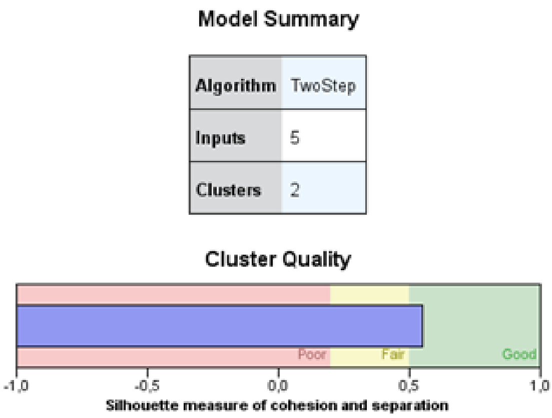

Fig. 19. Cluster model summary

Cluster 1 “No wind, no drugs, lighting”: This cluster mainly consists of cases of all countries except for Netherlands (UK, France, Italy, Greece and Poland). Moreover, 327 out of 398 cases concern no windy conditions at all. Similarly, 384 cases involve no drug involvement. Despite the fact that 368 cases out of 398 regard no alcohol involvement, it is interesting that almost all alcohol involvement cases are also included in that group. Finally, the vast majority of cases of this group include some type of lighting conditions, either natural (daylight, twilight) or artificial (electric light).

Cluster 2 “Windy, lighting, unknown DUI condition”: This cluster mainly consists of Netherlands and some UK cases. Moreover, 55 out of 99 cases concern mild or strong windy conditions. However, no information is known about DUI conditions (alcohol or drugs). Finally, the vast majority of cases of this group include some type of lighting conditions, either natural (daylight, twilight) or artificial (electric light).

Random Forest Analysis

Figure 20 shows the final results of variable importance rankings when examining injury severity (Fatal vs non-Fatal). It was produced by the method of Random Forests, as described in the Methodology section. The variable importance as unveiled from the Random Forest models are a helpful indicator to define which variables are significant. However, the magnitude of the effect and the sign of each variable are not identified.

It should be noted that variable importance should be interpreted as a relative ranking of predictors, since the absolute values of the importance scores should not be interpreted or compared over different studies, and furthermore statistically non-significant variables have values hovering around zero [25-26]. Variables to the right of dashed red vertical line are identified to be significant in an ascending order. This red vertical line on the plot is set at the value of the lowest important variable. Therefore, it can be observed that the most important factors (variables) are speeding, driving above speed limit and medication followed by use of narcotics.

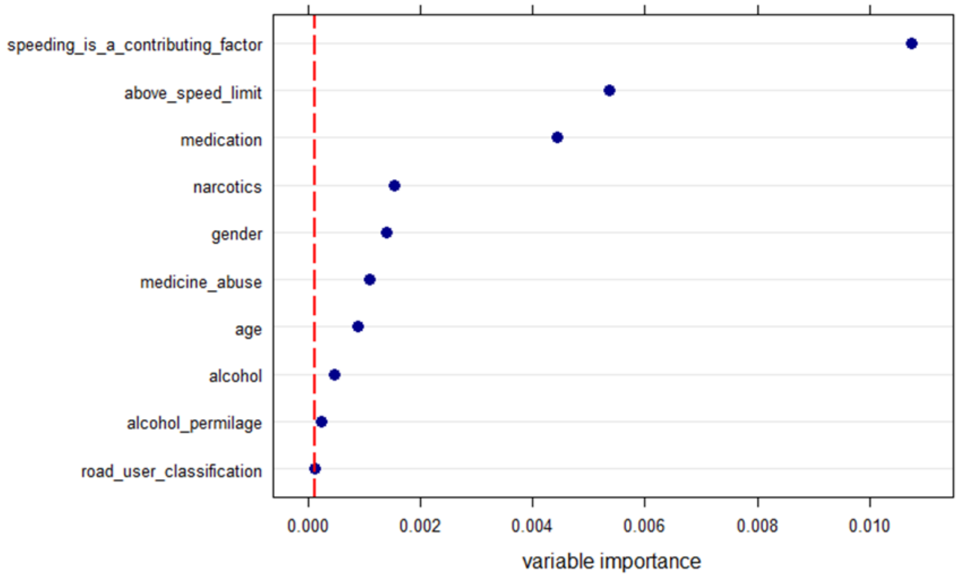

Fig. 20. Variable importance ranking from Random Forest analysis

IV. DISCUSSION

The previous results can offer some useful insights when viewed collectively.

It appears that regardless of origin, two-wheeler occupants remain vulnerable road users, with a considerable amount of serious and fatal injury accidents, which indicates that more severe accidents are more commonly and systematically collected for two-wheelers (especially in countries using retrospective methods).

There seem to be more active periods of time in which more accidents occur. For two-wheelers, these are the months between May and September, which are generally periods of good weather for the countries examined. Furthermore, in general tourism increases during that period, especially for southern European countries such as Greece and Italy. Additionally, more accidents seem to be caused during weekdays as opposed to weekends, and more accidents seem to be caused during working hours than non-working (off-peak) hours. It can be thus assumed that more active periods with increased two-wheeler trips (and thus exposure) lead to more accidents.

Regarding the average profiles of riders that are involved in accidents, PTW drivers that speed seem to be more often younger, male people overall. Perhaps this is explained via the criteria of each age group: speed, manoeuvrability and sensation seeking can be said to be the needs of younger people. Conversely, elder individuals might seek slower travelling speeds the comfort of a car, switch to a bicycle or on foot, or limit their exposure altogether (by travelling in fewer trips).

In line with the current scientific literature, younger age groups displayed higher instances of speeding during the accident, and speeding in turn led to more serious injuries in accidents (more fatal/serious injury accidents compared to slight/no injury accidents). Lastly, when riding a high-powered PTW, speeding was found to cause accidents more frequently.

As for the usage of protective equipment, most two-wheeler riders recognise the essentiality of helmet use while riding. The same cannot be said for reflective clothing. Helmets were also found to stay on during the accident. For the aspect of conspicuity, headlights were also used by the majority (81%) of PTW riders.

While there was considerable diversity in the PTW fleet (several different motor displacement categories were recorded), the overall condition of that fleet can be said to be good or excellent (85% in total). Furthermore, in very few instances were mechanical problems explicitly recorded in the vehicles, thus hinting that vehicle problems are not primary factors of accident occurrence.

When looking at accident circumstances, it was found that the highest amount of accidents was recorded in residential and commercial areas, during daylight conditions, in good weather and dry surface conditions and in local or collector roads. Again, this is explained via exposure, as these conditions are the more favourable ones for two-wheeler trips. The majority of accidents happen within areas with a speed limit of a 50km/h followed by 30 km/h, again indicating that two-wheelers are favoured for more urban uses. A noteworthy find is the very low accident numbers reported in roundabouts, as opposed to crossroads, T or Y junctions, and most importantly, single roads.

It is noted that the interpretation these conclusions should take into account the lack of the respective exposure data (vehicle kilometres per each user, road, traffic and vehicle type). Collection of these types of data was outside of the scope of the in-depth analysis; nevertheless it provides a strong direction for future research steps.

This study is highly relevant to research regarding injuries and their prevention, and to determine the causes leading to severe injuries. It has been found in the literature that the risks of functional, psychological and socio-economic consequences increase as a function of the injury severity, although the exact severity of injuries appears to have only little impact on psychological consequences. In addition, as less severe injuries are much more common than severe ones, they are responsible for a disproportionately high percentage of disabilities [27]. In addition, road serious injuries are treated as a road safety indicator; the determination of their numbers and other characteristics is starting to become more sophisticated in modernized countries, though further harmonization is needed [28]. It is apparent that in-depth analyses provide much needed light towards that direction.

V. CONCLUSIONS

The findings of the Two Step Cluster Analysis can be considered as a preliminary analysis of the in-depth accident database and can be used as a first guidance to provide insight on in-depth accident cases of powered to wheeler and bicycles as they revealed useful accident patterns. These results and the formation of the two clusters are as follows: the first cluster includes more favourable conditions ("no wind, no drugs, good lighting") while the second cluster consists of less favourable conditions for road safety ("windy, lighting, unknown DUI condition"). This hints at a meaningful separation of the examination of two-wheeler accidents when the influence of outside factors is considerable.

From the in-depth analysis several key points were determined: Several factors seem to affect accidents, and most of them appear to originate from road user behaviour, either actively to the accident location (such as speeding relatively to the road segment) or preceding the accident (substance consumption). Two-wheeler accidents with high road user mortality also appear to increase under specific factor combinations. The current study actively contributes to current knowledge by providing insights for factors affecting powered two-wheeler and bicycle accidents in practice. The inclusion of different but representative areas across Europe offers robustness and transferability to the data and respective results.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the colleagues of the respective research teams from collaborating institutes (Loughborough University, UK; SWOV, The Netherlands; CEESAR, France; CERTH-HIT, Greece; CTL, Italy; ITS, Poland; NTUA, Greece) that undertook the data collection and processing for the formation of the dataset. This research was funded by the European Commission.

VII. REFERENCES

[1] Adminaite, D. Jost, G., Stipdonk, G., & H & Ward, H. (2017). Ranking EU progress on road safety: 11th road safety Performance Index (PIN) report.

[2] World Health Organization – WHO. (2015). Global status report on road safety 2015.

[3] Vanlaar, W., & Yannis, G. (2006). Perception of road accident causes. Accident Analysis and Prevention, 38, 155–161.

[4] Thomas, P., Morris, A., Talbot, R., & Fagerlind, H. (2013). Identifying the causes of road crashes in Europe. Annals of Advances in Automotive Medicine, 57, 13–2
[5] Pai, C. W., & Saleh, W. (2008). Modelling motorcyclist injury severity by various accident types at T-junctions in the UK. Safety Science, 46(8), 1234–1247.
[6] Albalate, D., & Fernandez-Villadangos, L. (2010). Motorcycle injury severity in Barcelona: the role of vehicle type and congestion. Traffic injury prevention, 11(6), 623–631.
[7] Shankar, V., & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of safety research, 27(3), 183–194.
[8] Shaheed, M. S. B., Zhang, W., Gkritza, K., & Hans, Z. (2011). Differences in Motorcycle Conspicuity-related Factors and Motorcycle Crash Severities in Daylight and Dark Conditions. In 3rd International Conference on Road Safety and Simulation.
[9] Shaheed, M. S. B., Zhang, W., Gkritza, K., & Hans, Z. (2013). A mixed logit analysis of two-vehicle crash severities involvinga motorcycle. Accident Analysis and Prevention, 2013, 119–128.
[10] Shaheed, M. S. B. & Gkritza, K. (2014). A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research, 30–38.
[11] Pai, C. W., & Saleh, W. (2007). An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. Safety science, 45(8), 832–847.
[12] Chang, L.-Y., Wang, H.W. (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention, 38, 1019–1027.
[13] Vlahogianni, E. I., Yannis, G., & Golias, J. C. (2012). Overview of critical risk factors in Power-Two-Wheeler safety. Accident Analysis & Prevention, 49, 12–22.
[14] Theofilatos, A., Yannis, G., Vlahogianni, E. I., & Golias, J. C. (2017). Stochastic Cusp Catastrophe Models with Traffic and Weather Data for Crash Severity Analysis on Urban Arterials (No. 17-02516).
[15] DaCoTA project website (information, overview): http://safetyknowsys.swov.nl/Home/about.html
[16] Hill, J., Aldah, M., Talbot, R., Martinsson, J., Fagerlind, H. (Eds) (2012). Final Updated Protocol with Updates from the Pilot Review, Deliverable 2.4 of the EC FP7 project DaCoTA.
Available at: http://www.dacota-project.eu/deliverables.html
[17] Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.
[18] Strobl, C., & Zeileis, A. (2008). Danger: High power! — Exploring the statistical properties of a test for random forest variable importance. Proceedings of the 18th international conference on computational statistics, Porto, Portugal.
[19] Abdel-Aty,M., & Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accident Analysis and Prevention, 43, 461–470
[20] Ahmed, M., & Abdel-Aty, M. (2012). The viability of using automatic vehicle identification data for real-time crash prediction. IEEE Transactions on Intelligent Transportation Systems, 13(2), 459–468.
[21] Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science, 63, 50–56
[22] 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.
[23] Aarts, L., Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention, 38(2).
[24] Imprialou, M., Quddus, M., Pitfield, D., Lord, D. (2016). Re-visiting crash–speed relationships: A new perspective in crash modelling. Accident Analysis & Prevention, 86, 173-185.
[25] Strobl, C., Malley, J., Tutz, G. (2009a). An introduction to recursive partitioning: Rational, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14(4), 323-348.
[26] Strobl, C., Malley, J., Tutz, G. (2009b). Supplement to "An introduction to recursive partitioning: Rational, application, and characteristics of classification and Regression trees, bagging, and random forests." Retrieved from: http://dx.doi.org/10.1037/a0016973.supp.
[27] Weijermars, W. (2017), Impacts of Serious Road Injuries, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu on 22-03-2018.
[28] Pérez, K., Weijermars, W., Amoros, E., Bauer, R., Bos, N., Dupont, E., Filtness, A., Houwing, S., Johannsen, H., Leskovsek, B. Machata, K., Martin, JL., Nuyttens, N., Olabarria, M., Pascal, L., Van den Berghe, W., (2016), Practical guidelines for the registration and monitoring of serious traffic injuries, D7.1 of the H2020 project SafetyCube.