Investigating the correlation of mobile phone use with trip characteristics recorded through smartphone sensors

P. Papantoniou[1], A. Kontaxi[2], G. Yannis[3]

[1] National Technical University of Athens, 5 Heroon Polytechniou str., GR-15773, Athens, Greece, ppapant@central.ntua.gr
[2] National Technical University of Athens, 5 Heroon Polytechniou str., GR-15773, Athens, Greece, akontaxi@mail.ntua.gr
[3]National Technical University of Athens, 5 Heroon Polytechniou str., GR-15773, Athens, Greece, geyannis@central.ntua.gr

Abstract. The Internet of Things (IoT) constantly offers new opportunities and features to monitor and analyze driver behaviour through wide use of smartphones, effective data collection and Big Data analysis. In that framework, the aim of this paper is to investigate the impact of mobile phone use on driving behaviour and road safety through the investigation of driving analytics collected by smartphone sensors. For this purpose, a 100-driver naturalistic experiment was carried out and an innovative data collection scheme using a smartphone application was exploited in order to record the respective driving performance data. Then mixed binary logistic regression models were developed in order to investigate whether mobile phone use during a trip is correlated with driving metrics and can therefore be accurately forecasted based on them. Finally, a model for all trips was developed, as well as models for trips on different road types (urban, rural, highway). Exposure metrics found to be significantly associated with the probability of mobile phone use are total trip distance and driving on workdays and during rush hours. Additionally, the average speed is negatively associated with the probability of mobile phone use while driving.

Keywords: Mobile Phone, Driving Behaviour, Naturalistic Experiment, Smartphone Sensors, Mixed Binary Logistic Regression.

1 Background and Objectives

Road safety is a complicated scientific field of transport research due to the random nature of accident occurrence. Accidents impose serious problems to society in terms of human costs, economic costs, property damage costs and medical costs. There is a significant number of risk factors identified in literature, which affect accident probability. The most important risk factors recognized in literature (WHO, 2018; Elvik, 2004) are human factors (speeding, distracted driving, driving under the influence of alcohol and other psychoactive substances etc.), unsafe road infrastructure, unsafe vehicles and inadequate law enforcement of traffic laws. Among them, human factors are likely to be the most crucial cause of road traffic fatalities and injuries every year and therefore the importance of studying how these factors can affect traffic risk is high (Salmon et al., 2011).

Regarding distracted driving, mobile phone use (handheld or hands-free) and complex conversation (at mobile phone or with passengers) appear to be the most critical in-vehicle distraction factors (Papantoniou et al., 2015). Given that mobile usage is an inevitable part of everyday driving process and is expected to increase over the years (Stutts et al., 2001), its impact on driving behaviour in traffic and road safety is particularly important and should be further investigated. Literature so far has showed that when the driver is using the phone while driving, this is reflected to the behaviour expressed in terms of loss of control, response to incidents, or accident occurrence (Bellinger et al., 2009; Fitch et al., 2015; Dingus et al., 2016).

Although the importance of driver monitoring is established in the transportation field, researchers have been struggling with the difficulty of collecting accurate real-time driving data by adopting low-cost collection and processing methods. In that environment, the high penetration rate of smartphones and social networks nowadays provide new opportunities and features to monitor and analyze driver behaviour. The use of smartphones allow for drivers to be recorded under normal driving conditions and without any influence from external parameters, resulting at being considered as one of the most appropriate methods for the assessment of driving behaviour (Ziakopoulos et al., 2020).

Many studies have shown interesting results using data collected through smartphone sensors under naturalistic driving conditions. Researchers aim at examining the effect of various driving behaviour indicators on driver performance and cash risk (Tselentis et al., 2018, Papadimitriou et al., 2019, Paefgen et al., 2012) or at identifying aggressive and dangerous driving profiles through a clustering approach (Mantouka et al., 2018, Ma et al., 2017, Meseguer et al., 2013).

In light of the aforementioned, the aim of the current research is the utilization of high-resolution smartphone data for the investigation of the impact of detailed trip characteristics on the use of mobile phone while driving. To that end, a naturalistic driving experiment was carried out in order to model the probability of the mobile phone use while driving in all type of roads, and separately in each road type (urban, rural, highway).

2 Data collection

In order to achieve the research objective, an innovative smartphone application developed by OSeven Telematics (www.oseven.io) was exploited aiming to record driver behaviour using the hardware sensors of the smartphone device. One hundred and sixteen drivers participated in the related experiment during a 2-months timeframe and a large database of several thousand trips is created. The solid integration platform for collecting, transferring raw data and recognizing the driving behaviour metrics via ML algorithms is also developed by OSeven Telematics.

The standard procedure that is followed every time a new trip is recorded by the application is clearly presented in Figure 1. The data collected are highly disaggregated in terms of space and time. Once stored in the backend cloud server, they are converted into meaningful driving behaviour and safety indicators, using signal processing, Machine Learning (ML) algorithms, Data fusion and Big Data algorithms. Machine learning methods (filtering, clustering and classification methods) are mainly used to clean the data from noise and errors, and to identify repeated patterns within the data.

Fig. 1. The OSeven data flow system.

A variety of different metadata are eventually calculated, including indicatively the following exposure indicators:

- Total distance (mileage)
- Driving duration
- Type(s) of the road network used (given by GPS position and integration with map providers e.g. Google, OSM)
- Time of the day driving (rush hours, risky hours)

The driving behaviour indicators that are also calculated from the data include indicatively:

- Speeding (duration of speeding, speed limit exceedance etc.)
- Number and severity of harsh events
- Harsh braking (longitudinal acceleration)
- Harsh acceleration (longitudinal acceleration)
- Distraction from mobile phone use (mobile phone use is considered any type of phone use by the driver e.g. talking, texting etc.).

Finally, it must be noted that since privacy and security consist two critical principles in the field of telematics, the OSeven platform has very clear privacy policy statements for the end users covering the type of data collected, the reason data is collected for, the time that data is stored and the procedures for data security based on encryption standards for data in transit and at rest. All this is done using state-of-the-art technologies and procedures in compliance with standing Greek and European personal data protection legislation (GDPR).

3 Theoretical Background

In the present analysis, the mobile phone use while driving is examined and correlated with trip characteristics recorded through smartphone sensors. The examined variable can be available either as the percentage of time using the mobile phone per trip driving duration or as a binary regarding the entire trip in a form of yes/no aspect. Herein, the binary variable is selected for modelling the use of mobile phone.

Since, the dependent parameter consists of a binary variable, binary logistic regression is selected as the appropriate analysis method. A binary logistic regression estimates the probability that a characteristic is present (e.g. estimated probability of "success") given the values of explanatory variables; π=Pr (y=1|X=x). It leads to the development of a mathematical model that gives the odds of this event occurring, depending on factors that affect it. The odds are expressed by the logit link function as follows:

$$logit(\pi_i) = logit\frac{\pi_i}{1-\pi_i} = \beta_0 + \Sigma\beta_\iota x_\iota \quad (1)$$

The related outcome is specified as:

$$y_{i=}\ \beta_0 + \Sigma\beta_\iota x_\iota + e_{0i} \quad (2)$$

However, one may also consider that in the present dataset there are repeated measurements (trips) over the same units (drivers). Due to these repeated measurements, the observations are no longer independent. Unless accounted for, this dependency may affect the accuracy of the modelling results. In fact, it is necessary to account for random heterogeneity due to differences between drivers, so as to make sure that the effects identified in the model are true effects of the independent variables on the dependent, and do not reflect unobserved differences between drivers.

Therefore, in order to capture personal driver traits, such as personality and experience, which affects their driving style, random effects are introduced to Binary logistic regression Models in order to extend them as Mixed Binary logistic regression Models. In other words, Mixed Binary logistic regression Models are formulated by using the fixed effects Binary logistic regression Models as basis and then adding random effects to the equation. A mixed model (or random effects model, or multi-level model) is a standard technique in this context, i.e. a statistical model containing both fixed effects and random effects. The formulation of the linear mixed effects model, assuming a random intercept reflecting the repeated measurements (i) over drivers (j), is as follows:

$$y_{ij=}\ \beta_{0j} + \Sigma\beta_\iota x_{\iota j} + e_{0ij} \quad (3)$$

$$\beta_{0j} = \ \beta_0 + u_{0i} \quad (4)$$

It is noted that the intercept in the outcome Eq. (2b) consists of two terms: a fixed component β_0 and a driver-specific component, i.e. the random effect u_{0j} which is assumed to be normally distributed. The trip specific error term e_{0ij} in Eq. (2a) is assumed to follow a logistic distribution.

4 Results

4.1 Descriptive Statistics

Overall, during the 2-months timeframe of the experiment 11.987 trips from a sample of 100 car drivers have been recorded. However, for the present analysis a crucial criterion was set; all drivers chosen to be included in the analysis were required to have driven at least for 40 trips. This number approximately equals the typical monthly number of working trips for a driver assuming that each driver drives 2 trips a day for 5 working days a week. This number is reasonable to filter out drivers for which there are not enough observations, and it is also the 'industrial' criterion set by OSeven to start providing driver evaluation. As a result, from the 100 car drivers, 82 were ultimately selected creating a large dataset of 11.398 trips. While this might seem like a considerable loss of driver diversity, only 589 trips in total were removed from the initial dataset (resulting in a 4.91% reduction).

Before presenting the models development, explanatory descriptive analysis of the data is implemented, allowing for an overview of the examined risk factor; the use of the mobile phone while driving. In this context, the use of the mobile phone in this subsection is presented as a share of trip time during which mobile phone was used (% of trip duration). A first interesting observation is the different frequency of mobile use with respect to the type of road network. Specifically, it is found that the share of mobile use is more frequent on urban and rural roads in comparison with highways, which seems logical, taking into consideration the driving pattern in highways; longer distances covered and high speeds developed (figure 2).

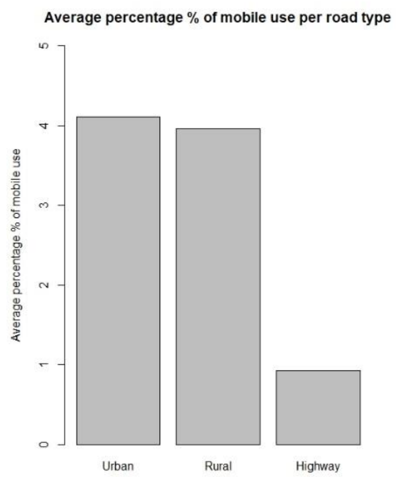

Fig. 2. Average percentage of mobile phone use per road type

Furthermore, figure 3 illustrates the average percentage of mobile use of the sample collected on a driver basis (left) and on a trip basis (right). It is evident that the majority of drivers show an average of 2% mobile phone use during their trips, which is reflected in the histogram developed for the number of trips, as well. However, it should be noted that there are a few trips where the share of mobile phone use reaches a maximum value equal to 60% of the total trip duration.

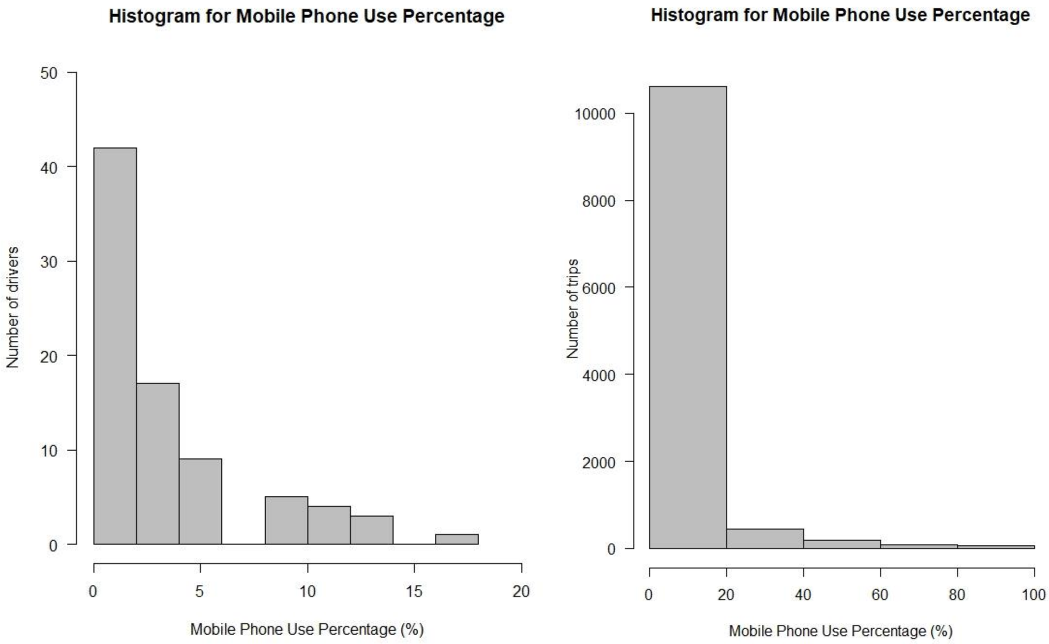

Fig. 3. Histogram of the percentage of time using the mobile phone per trip driving duration in of the driving sample's per driver (left) and trip characteristics (right)

Another interesting finding, concerns the correlation of mobile phone use and harsh events recorded during a trip duration. Harsh acceleration, harsh braking and harsh cornering events can be considered highly significant indicators that reveal an aggressive and unsafe driving behaviour (Vaiana et al., 2014). Fig. 4 presents a scatter plot illustrating the average number of harsh brakings (left) and harsh accelerations (right) per trip duration, against the share of mobile phone use during the trip on a driver basis. Overall, the scatter plots reveal a nonlinear relationship between the two driving behaviour indicators for both event types, resulting in a similar pattern despite the difference in the average number of harsh brakings and harsh accelerations, respectively. More specifically, harsh events are slightly less frequent when mobile phone is used either for a small share of the trip, or for the largest share of the trip.

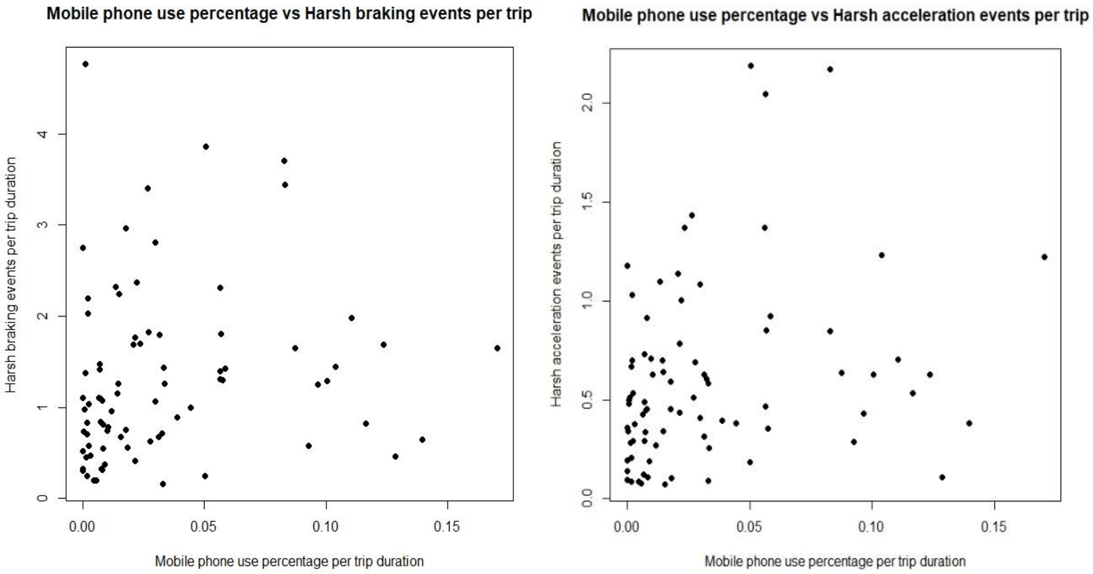

Fig. 4. Percentage of time using the mobile phone per trip driving duration in relation with the average of harsh braking (left) and harsh acceleration (right) events per trip duration

4.2 Mixed Binary Regression Models

In order to model the expected use of mobile phone while driving, models in a Mixed Binary Regression framework were calibrated, as previously explained. Since the OSeven application allows for a high resolution, big-data oriented collection scheme, it was attempted to include random effects in order to capture the unique driving behaviour traits for each driver. This entails having a critical minimum sample of trips for each driver to achieve a meaningful outcome. Therefore, a screening was made among participant drivers, as described above, and drivers that had over 40 trips each were selected for the analysis.

Mixed Binary Regression Models were fitted in R-studio (with the lme4 package) via maximum likelihood. A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. The fixed effects are attributed to the explanatory variables and the random effect is attributed to the model intercept. An overall model is developed for all road types, as well as separate models for each road type: urban, rural and highway.

The selected variables were chosen after taking into account the following: high statistical significance, low correlation among them, and final rational interpretation of their impact on the dependent variable. Due to the last criteria, authors did not include harsh events in the model, as harsh events would be the result of a distracted driving and not the opposite. Table 1 provides a description of the parameters that were selected for models.

Table 1. Description of the parameters used in the models.

Independent Variable	Description
Trip Distance	Total trip distance (km)
Workday	Workday (0=yes, 1=no)
Morning Rush	Morning rush hour 06:00-10:00 (0=yes, 1=no)
Afternoon Rush	Afternoon rush hour 16:00-20:00 (0=yes, 1=no)
Average Speed	Average driving speed per trip (km/h)

The final models are presented in Table 2. The sign of "–" in the table indicates that the specific variable was not used in the particular model. In all models, the variance of the random intercept is statistically significant; indicating that part of the variation is indeed due to unobserved differences between drivers.

Table 2. Mixed binary logistic models for all road types, and separately for urban roads, rural roads and highways

Parameter	Overall Model		Urban Road		Rural road		Highways	
	B	P-value	B	P-Value	B	P-value	B	P-value
Intercept	-1.613	<0.001	-2.313	<0.001	-2.752	<0.001	-6.457	<0.001
Trip Distance	0.051	<0.001	0.182	<0.001	0.095	<0.001	0.025	<0.001
Workday	0.174	0.003	0.176	0.005	0.174	0.008	-	-
Morning Rush	-0.354	<0.001	-0.385	<0.001	-0.44	<0.001	-0.704	<0.001
Afternoon Rush	-	-	0.121	0.046	0.127	0.045	-	-
Average Speed	-0.010	<0.001	-0.007	0.017	0.008	<0.001	0.037	<0.001
Random effect (variance of random intercept)	1.475	<0.001	1.515	<0.001	1.472	<0.001	1.763	<0.001
Number of obs	11398		11398		11398		11398	
Number of drivers	82		82		82		82	
AIC	10888.3		9517.6		9002.0		1637.7	

Modelling results regarding the use of mobile phone while driving (yes/no) reveal interesting findings; the parameters of trip distance, workday and afternoon rush have all been determined as statistically significant and positively correlated with the use of mobile phone. In the same context, average speed and morning rush are statistically significant and negatively correlated with the use of mobile phone.

More specifically, the aforementioned results could be further interpreted, calculating the relative risk ratio of every variable and thus measuring the increase in log-odds of mobile phone use. The exposure metric of trip distance seem to increase the odds of mobile phone use during the trip; the effect appears to be higher in urban areas, less in rural and the least in highways, as the respective parameters (B) are 0.182, 0.095 and 0.025, corresponding to odds ratios exp (B) of 1.20, 1.10, and 1.02. Driving on workdays compared to driving on weekends also increases the odds of mobile phone use, for all the developed models apart from the highway model, where the variable is not statistically significant. This impact may be interpreted by the fact that the busy daily schedule make drivers use more often their mobile phone than on weekends, when the purpose of the trip is not work-related.

In the same context, rush hours have also a significant impact on mobile phone use. Particularly, it is found that driving during morning rush hours (06:00-10:00) compared to the rest of the day decreases the odds of using the mobile phone during the trip. This can be explained by the fact that very early in the morning there are less mobile work-related calls. The impact of the morning rush hours appears to be higher in highways than in the other types of road, as well as the overall model. On the contrary, driving during afternoon rush hours (16:00-20:00) increases the odds of using the mobile phone while driving. The effect is significant only in model of urban and rural roads.

Finally, regarding the driving behaviour indicator of average speed, some conflicting results are revealed. Average speed per trip was found to be negatively associated with the odds of mobile phone use on all road types and in urban areas (odds ratios exp(B =−0.010) and exp(B=-0.007) equal to 0.99 respectively) as drivers adapt compensatory behaviour, confirming existing studies. However, when driving in rural areas and highways, it seems that the higher the average speed the higher the odds of mobile phone use. This is probably explained by the fact that, while in urban area the driver is aware of unexpected incidents that occur often and reduces the speed while talking on the mobile phone, outside urban areas, where the events occur less often, the compensatory behaviour of the drivers weakens and the speed is higher.

5 Conclusions and Future Research

This present research aimed to investigate the possibility of detecting risky driving behaviour on the basis of driver exposure and behaviour metrics collected by smartphone sensors, focusing on the use of mobile phone while driving. In order to achieve that objective, a naturalistic driving experiment was carried out in order to examine distracted driving as expressed by the use of mobile phone while driving. Results reveal correlations of mobile phone use with specific driving behaviour and exposure metrics, namely the average driving speed, the trip distance as well as driving on workdays and rush hours.

Furthermore, the present study does not only investigate the mobile phone use, but also allows for mobile detection while driving by means of mobile sensors. This is very essential, taking into consideration the high penetration rate of smartphone use in our everyday life. Future research should go one step further and focus on real-time detection of mobile phone use providing new possibilities to driver feedback, aiming at the reduction of mobile phone use while driving and therefore improving the overall driving behaviour.

Future research should also focus on the improvement of the accuracy of the models, by exploring more variables and alternative modelling techniques. Furthermore, with respect to mobile phone use, analyses per gender, age, history of accidents, self–assessment, driving experience and more demographic characteristics could be undertaken in order to capture any particular trends found in the categories of these parameters.

Acknowledgements

This research is based in the DistrApp project which is co-financed by Greece and European Union (Social European Fund) through the Operational Program "Human Resource Development, Education and Lifelong Learning", within the action "Strengthen Postdoctoral Researchers - Second Phase" (MIS 5033021), implemented by the State Scholarships Foundation (IKY)

The authors would like to thank OSeven Telematics, London, UK for providing all necessary data exploited to accomplish this study.

References

1. Bellinger, D.B., Budde, B.M., Machida, M., Richardson, G.B., Berg, W.P. (2009). The effect of cellular telephone conversation and music listening on response time in braking. Transportation Research Part F 12 (6), pp. 441-451.
2. Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proceedings of the National Academy of Sciences, 113(10), 2636-2641.
3. Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and road accidents. An evaluation of the Power Model. TØI report, 740, 2004.
4. Fitch, G. M., Bartholomew, P. R., Hanowski, R. J., & Perez, M. A. (2015). Drivers' visual behaviour when using handheld and hands-free cell phones. Journal of safety research, 54, 105-e29.
5. Ma, C., Dai, X., Zhu, J., Liu, N., Sun, H., & Liu, M. (2017). Driving sense: Dangerous driving behavior identification based on smartphone autocalibration. Mobile Information Systems, 2017.
6. Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2018). Mobile sensing and machine learning for identifying driving safety profiles (No. 18-01416)
7. Meseguer, J. E., Calafate, C. T., Cano, J. C., & Manzoni, P. (2013). Driving styles: A smartphone application to assess driver behavior. In 2013 IEEE Symposium on Computers and Communications (ISCC) (pp. 000535-000540). IEEE.
8. Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012). Driving behavior analysis with smartphones: insights from a controlled field study. In Proceedings of the 11th International Conference on mobile and ubiquitous multimedia (pp. 1-8).
9. Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. Safety Science, 119, 91-97.
10. Papantoniou, P., Papadimitriou, E., Yannis, G. (2015). Assessment of driving simulator studies on driver distraction. Advances in Transportation Studies, 35, 129-144.
11. Salmon, P., Young, K., Lenné, M., Williamson, A., Tomasevic, N., 2011. The Nature of Errors made by Drivers. Austroads Publication No. AP–R378/11. Austroads Ltd., Australia.
12. Stutts, J.C., Reinfurt D.W., Staplin L., Rodgman E.A. (2001). The role of driver distraction in traffic crashes. Report Prepared for AAA Foundation for Traffic Safety.
13. Tselentis, D. I., Vlahogianni, E. I., & Yannis, G. (2018). Comparative Evaluation of Driving Efficiency Using Smartphone Data. *METHODOLOGY*, *32*, 33.

14. Vaiana, R., T. Iuele, V. Astarita, M. V. Caruso, A. Tassitani, C. Zaffino, and V. P. Giofrè. Driving Behavior and Traffic Safety: An Acceleration-Based Safety Evaluation Procedure for Smartphones. Modern Applied Science. 2014 Feb 1;8(1):88.
15. World Health Organization. Global Satus Report on Road Safety. 2018.
16. Ziakopoulos, A., Tselentis, D., Kontaxi, A., & Yannis, G. (2020). A critical overview of driver recording tools. *Journal of Safety Research*.