Correlation of declared and revealed driver behaviour using smartphone sensors

Konstantina Kokkali[1], Apostolos Ziakopoulos[1], Armira Kontaxi[1], George Yannis[1]

Department of Transportation Planning and Engineering – National Technical University of Athens, 5 Heroon Polytechniou Str., GR-15773 Athens, Greece

Email: nandiakokkali@gmail.com
Email: apziak@central.ntua.gr
Email: akontaxi@mail.ntua.gr
Email: geyannis@central.ntua.gr

Abstract

The aim of this study is the correlation of declared and revealed driver's behaviour with the use of smartphone sensors at a naturalistic driving experiment. To this purpose, data were collected from an innovative application for smartphones concerning harsh acceleration and braking, average speed, mileage travelled, etc. OSeven Telematics smartphone application was used to monitor naturalistic driving trips of 19 drivers, whereas drivers' stated behaviour was investigated through a related questionnaire. Data analysis was carried out through two regression Poisson statistical models: one model for harsh accelerations and one model for harsh breakings. Both models' results indicate that driving experience, driver age, number of injury accidents in which the driver was involved, vehicle age, fines received by the driver during the last 3 years are all variables associated with increased harsh events. In most of the cases, a convergence between stated and revealed behaviour was observed.

Keywords: road safety, driving behavior, declared and revealed behaviour, smartphone sensors, harsh acceleration, Generalized Linear Models

1. Introduction & State-of-the-art

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. According to World Health Organization (WHO), the total number of road fatalities worldwide continues to climb, reaching a high of 1.35 million in 2018. Regarding European Union, traffic accidents were the fifth cause of death in 2018, with roughly six people out of every 100,000 dying on the roads of the European Union because of road crashes. Consequently, understanding the various risk factors that cause road accidents is very crucial and has attracted great attention in the literature. Although there has been a very considerable research effort so far, there is still much to be investigated, especially in order to acquire a better knowledge of detailed pre-accident conditions in order to have a better proactive safety management in major roads of the transport network.

In a very short amount of time, with the evolution of technology, the automotive telematics market is growing steadily and a few innovative telematics and driver monitoring systems are introduced in our life. Nowadays, most drivers look for new services providing more options in order to identify their weak points in driving, adjust their driving style and techniques, reward their progress and promote maximum road safety for everyone and minimum motor vehicle emissions. Automotive telematics technology receive information from vehicles, including GPS coordinates, engine diagnostics, sensors, wireless internet connections, radar, touch screens as well as cameras inside and outside the vehicle and send it to a centralized server where it is then analyzed and managed using fleet management software. In this paper, studies that refer to the driver behaviour through technological devices adapted to the car's brain are presented. Many surveys were conducted to analyze driving behaviour characteristics using data from smartphones with focus on key risk indicators, namely the number of harsh driving events.

More specifically, in many studies that have taken place internationally, a device agnostic platform has been developed with the ability to collect data from different sources such as smartphones, OBDs and 4G/5G connected cars. Additionally, it appears that texting leads to statistically significant decrease of the mean speed and increase of the mean reaction time in urban and rural road environment. Simultaneously, it leads to an increased accident probability due to driver distraction and delayed reaction at the moment of the incident. It appeared that drivers using mobile phones with a touch screen alter their driving behavior with respect to their mean speed, however, they had an even higher probability of being involved in an accident (Yannis et al., 2014). Furthermore, aggressive driving, a particular type of driving style, has long been studied due to its strong correlation with accidents and traffic safety hazards: by one estimate, it was influential in causing the majority of accidents in the United States from 2003 to 2007 (Hong et al., 2014). Moreover, real driving parameters of driver behaviour have been assessed and analyzed through an On Board Diagnostics (OBD-II) device (Yannis et al., 2016). At the same time, in another study a methodology for evaluating driving behaviour by means of an Android application was developed (Vaiana et al., 2014). A similar survey indicates that excessive speed remains the number one causal factor associated with serious accidents in New South Wales. Exploratory analysis shows that speeding is more prevalent in high (100-110 km/h) and low (40-50 km/h) speed zones, and tends to be higher on weekday mornings and weekend nights. Overall, males seem to speed more than females but there are only marginal differences by age (Ellison et al., 2010).

Particular emphasis is put on research that focus on the correlation of declared and revealed driver's behaviour with the use of smartphone sensors. At first, OBD-II recording system was developed in the United States of America and it is designed to detect road accidents and mechanical problems in the vehicle. An android smartphone connects via Bluetooth to the OBD-II and receives information about the vehicle status, such as speed, fuel, temperature, accelerometer values as well as accurate location with a specific latitude and longitude, via GPS updates (Zaldivar et al., 2011). In addition, TrafficView defines a framework to disseminate and gather information about the vehicles on the road. The Ratio-based algorithm is more flexible than the other algorithms as it provides a better control over the tradeoff between the accuracy and visibility governed by the parameter settings. Regarding the other methods, although tuning the parameters is easier, the cost function does not provide the flexibility present in the Ratio-based algorithm (Nadeem et al., 2004). Another study explores the potential uses of feedback systems in the trucking industry as a means of improving safety. Since truck drivers spend the majority of their working time alone and do not interact with peers, it may be possible to use data gathered by in-vehicle technology to provide feedback to drivers about their driving behavior. However, most drivers were willing to accept feedback by technology if the program was designed properly. The truck drivers expressed no strong preference regarding the best form of feedback by technology on driving performance (Huang et al., 2005). Moreover, a driving behaviour and safety evaluation was conducted through a data recording system called Drive Diagnostics which is a dedicated In-Vehicle Data Recorder (IVDR) system with dimensions of 11x6x3cm, customized inside the vehicle (Toledo et al., 2006).

Based on the above, the objective of this paper is to investigate the correlation of declared and revealed driver behaviour with the use of smartphone sensors. To that end, a naturalistic driving experiment was carried out in order to examine whether driving characteristics recorded by smartphone sensors have a strong correlation with revealed driver behaviour.

2. Methodology

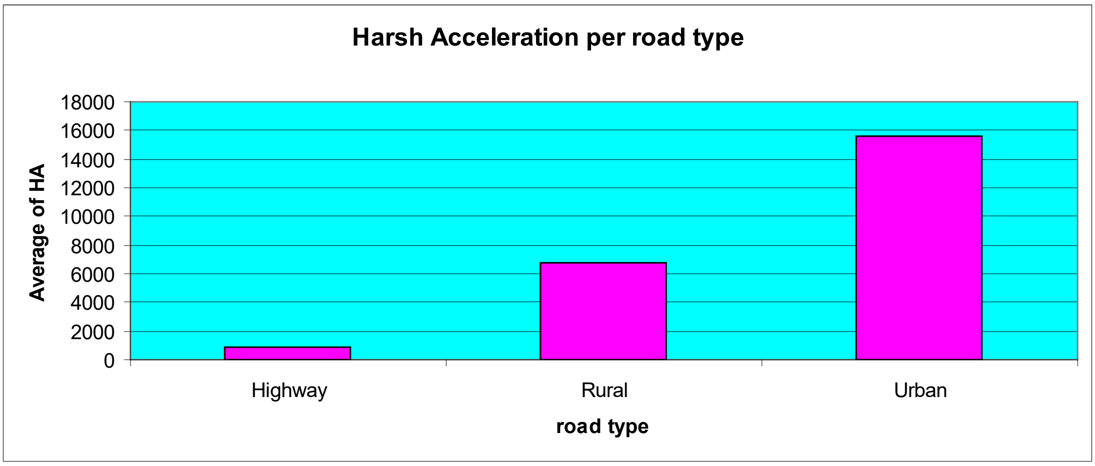
For the purpose of this research, a naturalistic driving experiment was carried out involving 19 drivers aged 28-43 and a large database of thousands trips was created. Driving behaviour analytics were recorded in real time, using smartphone device sensors. More specifically, an innovative data collection system using a Smartphone Application that has been developed by OSeven Telematics was exploited. A set of sophisticated and personalized interactive tools is applied by OSeven Telematics, powered by breakthrough technology, smart algorithms and reliable metrics. The most advanced Machine Learning techniques have been implemented to detect harsh events and speeding, identify the trip transport mode, and recognize whether the user is a driver or a passenger. Consequently, the OSeven platform helps drivers understand their weak points and motivates them to improve their driving behaviour. The standard procedure that is followed every time a new trip is recorded by the App, is clearly presented in Figure 1.

For every trip a driver completes, a large amount of data is recorded, transmitted through WiFi or cellular network and valuable critical information such as metrics, features, highlights and driving score is produced in order to evaluate driving profile and performance. The exposure indicators that are available include indicatively duration (seconds), total distance (mileage), type(s) of the road network used, given by GPS position and integration with map providers e.g. Google, OSM, (highway, rural or urban environment) time of the day driving (rush hours, risky hours) and weather conditions. Moreover, the driving indicators which can reliably quantify the risk associated with a specific driving behaviour are the following: speeding (distance and time of driving over the speed limit and the exceedance of the speed limit), driver distraction (caused by smartphone use during driving), number and severity of harsh events number and severity of harsh events (braking and acceleration), harsh cornerings and driving aggressiveness (e.g. braking, acceleration).

It should be noted that since privacy and security consist two of the main 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 GDPR. In this framework all data has been provided by OSeven Telematics in an anonymized format.

Before the driving experiment, each participant was requested to fill in a questionnaire which was divided into three distinct sections: a) overall driving data, b) attitude and behaviour towards road safety, and c) demographic characteristics. Specifically, drivers have provided valuable information such as age, gender, educational level, history of accidents, self–assessment, driving experience, vehicle type (cubism, fuel, etc.). They have also provided information on their driving behaviour, namely speed limits, traffic violations, harsh events (braking or/and acceleration) and mobile phone usage during driving.

The following chart constitutes a preliminary analysis of the variables, which allows for an initial better understanding of the data and the results and will be used to draw qualitative conclusions. Figure 2 illustrates the average harsh acceleration counts during driving for each different road type respectively. It is noted that the average of harsh accelerations is higher in the urban environment while on highways drivers realize harsh accelerations the least, probably because the harsh events on highways are not a common phenomenon, due to the high speeds developed in that type of road.



As previously reported, data from the driving measures collected by OSeven backend office and the questionnaire were analysed using Microsoft Excel and SPSS. Afterwards, statistical analyses were carried out using multiple Poisson models. This type of analysis was developed to examine whether driving characteristics such as speed, harsh acceleration, harsh braking, smartphone usage (dialing, talking, texting etc.) have a correlation of declared and revealed driver behaviour.

The general form of the GLM models the log odds via a linear predictor. Following McCulloch (2003), if y is the observed count of harsh events per trip i, and λ is the expected count of harsh events to be predicted, then the Poisson model is specified as:

$$y_i \sim Poisson\ (\lambda_i) \tag{1}$$

And the linear predictor is:

$$\log(\lambda_i) = \beta_o + \beta_n + x_n + \varepsilon \tag{2}$$

Where β are the fixed-effect parameters (constant and coefficients) for n independent variables, and ε is the error term.

In this study, two statistical models forecasting harsh events were developed: one for harsh accelerations and a second one for harsh brakings as shown below:

- Model 1: Predicting the frequency of harsh accelerations in all types of road
- Model 2: Predicting the frequency of harsh brakes in all types of road

Specifically, all models were designed to have exactly the same variables in order to make it easier to compare models with each other. Then, it was examined for all models separately whether the numerical results quantifying relationships between variables, did satisfy the models' quality. Within the present research, the scope of this analysis was to determine which observed independent variables are highly correlated with the dependent variable and at the same time which of them are inconsistent with each other. For this reason, the Pearson correlation coefficient, which indicates the percentage of the dependent variable that is explained by the independent variables, needs to be as low as possible.

It is important to mention that the values and signs of the Poisson coefficients bi must be reasonably explainable. Furthermore, the value of the statistical significance level should be acceptable and satisfactory for the confidence level commonly used. The constant coefficient of the equation, which indicates all the parameters that have not been taken into account, must be the lowest possible. Lastly, the elasticity (ei) that shows how responsive one variable is to a change in another but also the relevant influence elasticity (ei*) used for quantifying the influence of each individual variable should be examined allowing for the comparison between the influence of different variables in a single model.

3. Analysis & Results

The final dataset obtained from this study consisted of several types of variables regarding driver characteristics, parameters extracted from the naturalistic driving experiment as well as parameters extracted from the questionnaire. The dataset included both nominal and ordinal variables such as average speed, harsh acceleration and harsh braking, mobile phone use while driving, total distance, duration, gender, age etc. In the first model the dependent variable was the frequency of harsh accelerations and in the second one was the frequency of harsh break. The independent variables were the following: years_driving, age_vehicle, total accidents, cautious driver, age, job, vehicle_belongs, fines_last 3 years and skillful_driver. Table 1 provides a description of the independent variables that were found to be significant in the Poisson models.

Table 1: Description of the variables used in the analyses

Variables	Explanation
Years_driving	How many years the driver has been driving
Age_vehicle	The age of the vehicle
Total accidents	How many accidents the driver has been involved to, up to date?
Cautious driver	How cautious the driver think they are?
Age	Age of driver
Job	Profession of the driver
Vehicle_belongs	Who is the owner of the vehicle
Fines_last 3 years	How many driving fines the driver has received for road traffic violations during the last 3 years
Skillful_driver	How skillful the driver think they are

In Tables 2 and 3, the parameters of the driving experiment and the Poisson results are presented for harsh accelerations and harsh brakings, respectively. The developing models have given a more clear understanding of the correlation of declared and revealed driver behaviour, i.e. does the drivers' opinion inflect their driving behaviour.

Table 2: Poisson model outcomes correlating declared and revealed driver behaviour for harsh accelerations

Variables	**Model 1 – HA**			
	βi	ei	ei*	p-value
Years_driving	0.511	0.006	1.000	<0.001
Age_vehicle	-6.599	-0.452	38.818	<0.001
Total accidents	6.226	0.426	36.624	<0.001
Cautious driver	2.855	0.195	16.794	<0.001
Age	-0.522	-0.013	2.060	<0.001
Job	9.816	0.672	57.741	<0.001

Table 3: Poisson model outcomes correlating declared and revealed driver behaviour for harsh brakings

Variables	**Model 2 – HB**			
	βi	ei	ei*	p-value
Years_driving	0.283	0.009	1.000	<0.001
Total accidents	-1.002	-0.168	1.648	<0.001
Age	-0.182	-0.011	1.297	<0.001
Vehicle_belongs	-2.613	-0.437	4.298	<0.001
Fines_last 3 years	7.596	1.270	12.493	<0.001
Skillful_driver	-1.882	-0.315	3.095	<0.001

As shown in the first model, the driver's profession has the greatest influence to the dependent variable compared to the other independent variables. In the same context, the age of the vehicle was identified as the second most influential variable in the model. These results can be probably explained by the fact that routine and difficult everyday life force drivers to increase the speed and drive more nervously, pushing the accelerator pedal intensively. Moreover, the older the vehicle, the less the drivers tend to make harsh accelerations. Additionally, it was observed that drivers who have been involved in fewer accidents show less harsh accelerations while driving compared to the rest of the drivers. With respect to the other variables in the model, they seem to have a smaller impact on the frequency of harsh accelerations.

As far as Model 2 is concerned, drivers who have received fines for traffic code violations tend to make more harsh brakings while driving. This can easily be explained by the fact that those drivers may probably be afraid of being fined again for infringement. The next most influential variable is the "Vehicle_belongs"; when the vehicle is rented, the driver seems to drive in a calmer and more conservative way, in comparison with the drivers that own their vehicle and so illustrate an increased number of harsh brakings.

Furthermore, sensitivity analysis was conducted and is presented in Figure 3. The chart shows to what extent the increase in the driving experience influences harsh accelerations for different vehicle age groups. For this purpose, the values of all independent variables except from the driving experience remained fixed and the driving experience is plotted for these three vehicle age groups.

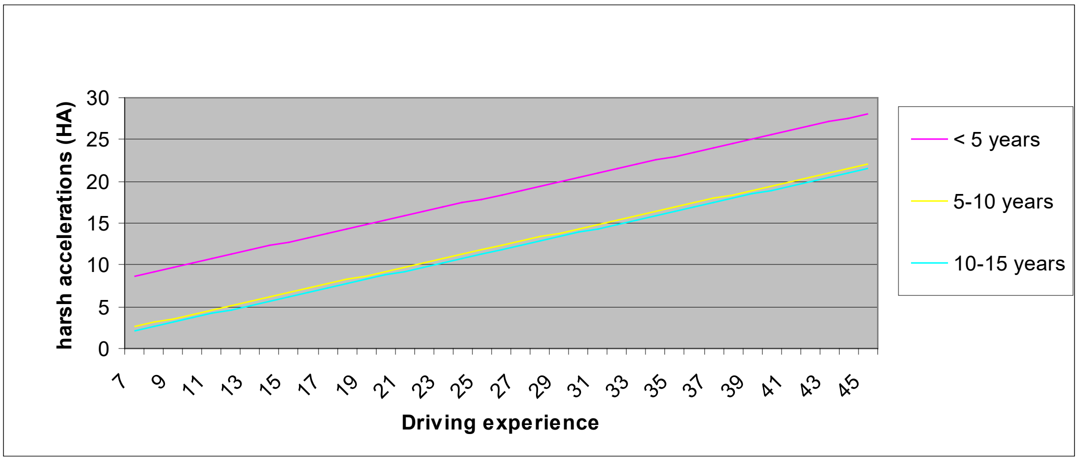


Figure 3. Correlation of harsh accelerations and driving experience for different vehicle age groups

Additionally, sensitivity analysis displayed in Figure 4 shows to what extent the increase in the driving experience, influences harsh brakings for the different means of vehicle ownership It is found that the higher the number of driving experience, the higher the number of harsh brakings occurred while driving.

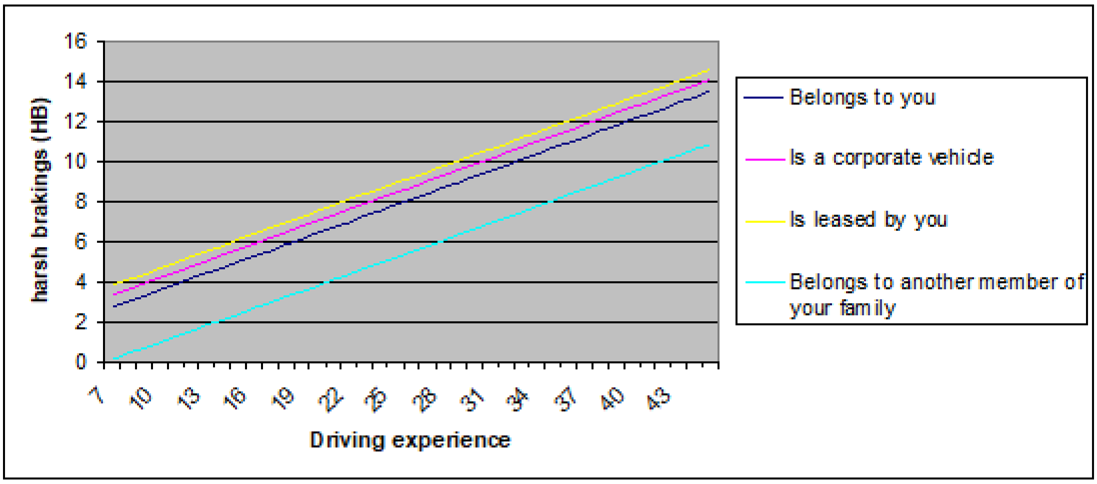


Figure 4. Correlation of harsh brakings and driving experience for different means of vehicle ownership

4. Conclusions & Future research

The aim of this study is the correlation of declared and revealed driver's behaviour through the use of smartphone sensors during a naturalistic driving experiment. To this purpose, data were collected from an innovative application for smartphones concerning harsh acceleration and braking, speeding, mobile phone use, mileage travelled etc. OSeven Telematics smartphone application was used to monitor naturalistic driving trips of 19 drivers, whereas drivers' stated behaviour was investigated through a related questionnaire. Results indicate that there is a strong correlation between harsh accelerations and driving experience. In addition, older drivers do not accelerate harshly as much as younger drivers. Furthermore, the age of the vehicle is a deterrent factor to the driver accelerating sharply, so it makes driving safer. Moreover, when the vehicle belongs to the driver or to the company where he works, the drivers make more harsh brakes. Also, drivers with lower fines for traffic code violations tend to make more harsh braking events.

At the same time, few innovative telematics systems are designed to specifically address these issues by working with the driver. It is expected that the technological devices will be the dominant player for the following years in the transport field as a fully scalable hardware-free solution. With the connected vehicles having been identified as the fastest-growing technological device after the smartphone and tablet, it gets easier to imagine the range of capabilities drivers can come to expect over the coming decade. The insights from the massive and highly-accurate driving behaviour data can play a crucial role to road safety by providing driver feedback; drivers can receive the latest reports on potentially dangerous behaviour, improve their weak points and test their driving skills. Moreover, the use of modern mobile devices like smartphones or tablets and their internal sensors such as GPS receivers and three axes accelerometers, allows road users to receive real time information on their behaviour that can be useful to increase awareness of drivers and promote safety.

Focusing on driver behaviour, drivers should understand that their style of driving significantly affects their whole life and if they have a correct idea about the way of their driving can bring benefits not only to themselves but also can safe thousands lives. One of the most fundamental tools to achieve a better and smoother driver behaviour is a change in traditional ideas, mentality and way of thinking. Undoubtedly, mindset is all about attitude which is also depicted through driving style of each one. Safer driving and therefore saving lives must be a strong behavioural change motivator that will help drivers to adopt a culture that puts safety as the first priority and takes it seriously. Drivers have to embrace driving habits that unlock their full potential as safe drivers, become more focused on the road, be more eager to drive safer and use clutch, brake and acceleration with the softer way. That is the most comprehensive approach to creating a safe driving culture.

Dozens of surveys reveal that recognizing risky driving behaviour can be a strong motivator for drivers to change their bad driving habits and improve their behaviour. It is self-evident that driving education is one of the most important aspects of the role of training. In particular, male drivers, who are over-represented in severe crashes compared to young female drivers, when also controlling for exposure (annual mileage), should be subjected in order to understand all these essential elements that can lead to a better and more safe driving behaviour. Also, driving training programs should enrich drivers with the proper driving skills and knowledge of the benefits of eco-driving, laws and rules of the roads and responsibilities of the driver so they can acquire a more prudent and safe behaviour and minimize their personal risks. Fully-educated drivers are more adept at handling their vehicle and learn to drive wisely even if they encounter in an unfamiliar traffic environment. Training results in 7-11% less crashes for novice drivers compared to untrained ones (SafetyCube, 2017).

Apart from the above, insurance companies could reward cautious drivers with lower insurance rates and costs over time and this can contribute to the reduction of road accidents and air pollutant gases. More specifically, usage-based motor insurance (UBI) schemes will play a key role in motor insurance market in the future and as a result it will strongly influence traffic safety in total. It is very essential to introduce economic incentives in form of a flexible discount on traffic charges depending on the driver's ecological and safe driving behaviour. Lastly, stricter police enforcement and more frequent harsher sanctions must be used to eliminate aggressive driving behaviour and reduce fuel consumption and emissions. Summarizing, driver training, in-vehicle technologies and innovative telematics systems constitute a key role by providing drivers information on efficient driving techniques, while the technologies provide ongoing reminders encouraging drivers to use and further develop their efficient driving skills.

The next steps of the present study include the organization of a following naturalistic driving experiment where a larger sample of drivers with different age groups could be carried out. The more drivers are participated at a naturalistic driving experiment, the more reliable the variables and results are. With regard to research methodologies, there are many different statistical data analysis methods, such as cluster analysis, factor analysis or time series analysis that can be used in order to icily exhibit different driving style at any types of road. Furthermore, it would be quite impressive to be achieved a similar experiment involving the worst-performing drivers of all ages.

Concluding, it is expected that this research can provide considerable gains to the society, since the stakeholders including policy makers and industry could rely on the results and recommendations regarding risk factors that appear to be critical for safe driving. As for further research, the examination of additional methods of analysis are proposed, such as factor analysis, logistic regression as well as microscopic data analysis of the database collected could be implemented through econometric techniques such as time-series analysis. Future studies would also benefit from exploiting more advanced technological equipment for recording the in-vehicle driving behavior such as precise GPS equipment, radars measuring the reaction time and headways as well as cameras inside and outside the vehicle. However, these solutions sometimes come at considerable costs, resulting in the investigation of affordable and ergonomic ways of monitoring and assessing driving behavior.

Acknowledgments

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

References

1. 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.
2. Ellison, A. B., & Greaves, S. (2010, September). Driver characteristics and speeding behaviour. In Proceedings of the 33rd Australasian Transport Research Forum (ATRF'10).
3. Elvik, R., Vaa, T., Hoye, A., & Sorensen, M. (Eds.). (2009). The handbook of road safety measures. Emerald Group Publishing.
4. Hong, J. H., Margines, B., & Dey, A. K. (2014). A smartphone-based sensing platform to model aggressive driving behaviors. In Proceedings of the sigchi conference on human factors in computing systems (pp. 4047-4056).
5. Huang, Y. H., Roetting, M., McDevitt, J. R., Melton, D., & Smith, G. S. (2005). Feedback by technology: Attitudes and opinions of truck drivers. Transportation Research Part F: Traffic Psychology and Behaviour, 8(4-5), 277-297.
6. McCulloch, C. E. (2003). Generalized linear mixed models. In NSF-CBMS regional conference series in probability and statistics (pp. i-84). Institute of Mathematical Statistics and the American Statistical Association.
7. Nadeem, T., Dashtinezhad, S., Liao, C., & Iftode, L. (2004). Trafficview: Traffic data dissemination using car-to-car communication. ACM SIGMOBILE Mobile Computing and Communications Review, 8(3), 6-19.
8. OSeven Telematics, (2019), https://www.oseven.io
9. Pavlou, D., Kyriakouli, E., Papantoniou, P., Yannis, G. (2019). Does driving at night affect the driving performance of young drivers? A driving simulator study. RSS 2019 conference, Iowa City, Iowa, USA.
10. Prato, C. G., Toledo, T., Lotan, T., & Taubman-Ben-Ari, O. (2010). Modeling the behavior of novice young drivers during the first year after licensure. Accident Analysis & Prevention, 42(2), 480-486.
11. Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C: Emerging Technologies, 16(3), 320-331.
12. Tselentis, D. I., Yannis, G., & Vlahogianni, E. I. (2017). Innovative motor insurance schemes: A review of current practices and emerging challenges. Accident Analysis & Prevention, 98, 139-148.
13. Vaiana, R., Iuele, T., Astarita, V., Caruso, M. V., Tassitani, A., Zaffino, C., & Giofrè, V. P. (2014). Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones. Modern Applied Science, 8(1), 88.
14. WHO (World Health Organization), Global Status Report on Road Safety (2019), www.nrso.ntua.gr/who-globalstatus- report-on-road-safety-highlights-insufficient-progress-2019
15. Yannis, G., Laiou, A., Papantoniou, P., & Christoforou, C. (2014). Impact of texting on young drivers' behavior and safety on urban and rural roads through a simulation experiment. Journal of safety research, 49, 25-e1.
16. Yannis, G., Papadimitriou, E., & Papantoniou, P. (2014). Distracted driving and mobile phone use: Overview of impacts and countermeasures. Commun. Technol. Road Saf, 1-23.

17. Yannis, G., Tselentis, D., Papadimitriou, E., & Mavromatis, S. (2016). Star rating driver traffic and safety behavior through OBD and smartphone data collection.
18. Zaldivar, J., Calafate, C. T., Cano, J. C., & Manzoni, P. (2011, October). Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In 2011 IEEE 36th Conference on Local Computer Networks (pp. 813-819). IEEE.