Analysis of driver behaviour through smartphone data
The case of mobile phone use while driving

Papadimitriou, Eleonora; Argyropoulou, Anastasia; Tselentis, Dimitrios I.; Yannis, George

DOI
10.1016/j.ssci.2019.05.059

Publication date
2019

Document Version
Accepted author manuscript

Published in
Safety Science

Citation (APA)
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. https://doi.org/10.1016/j.ssci.2019.05.059

Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving

Eleonora Papadimitriou[a1*], Anastasia Argyropoulou[b], Dimitrios I. Tselentis[b], George Yannis[b]

[a] Delft University of Technology, Faculty of Technology, Policy & Management, Safety & Security Science Section, Jaffalaan 5, 2628 BX Delft, The Netherlands

[b] National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou St., GR-15773 Athens, Greece

Abstract
The aim of this paper is to explore driving behaviour during mobile phone use on the basis of detailed driving analytics collected by smartphone sensors. The data came from a sample of one hundred drivers (18,850 trips) during a naturalistic driving experiment over four months. A specially developed smartphone application was used, through which driving exposure and behaviour metrics are captured by the smartphone sensors and transmitted to a back-end platform. The data are processed by Machine Learning algorithms yielding exposure (e.g. distance travelled per road type and time of day) and behaviour indicators (e.g. speeding, speed and acceleration variations, harsh braking, harsh manoeuvring, use of mobile phone etc.). Mixed binary logistic regression models were developed to investigate whether mobile phone use during a trip is correlated with other driving metrics, and can be accurately "detected" based on them. 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 trip length, and driving off-morning rush. Exceeding the speed limits and the number of harsh events (particularly harsh cornering), are all negatively associated with the probability of mobile phone use. A general pattern of less speeding and smoother driving appears indicative of mobile phone use, in line with known assumptions of driver compensatory behaviour. The results suggest that mobile phone use while driving may be accurately predicted by the model in more than 70% of cases.

Keywords: distraction, mobile phone use, smartphone sensors data, driver behaviour, road safety.

1. Introduction

1.1. Background

Mobile phone use while driving is persistently shown in the literature to have significant detrimental effects on driver behaviour and safety, due to the higher level of workload involved in such multi-tasking, regardless of the conversation difficulty level (Haque, 2015). Mobile phone use results in higher speed variation (Haque, 2015) and difficulties in maintaining vehicle lateral position (Horrey & Wickens, 2006). Lower driving speed is observed, and as a result increased vehicle headways, revealing a possible risk compensation behaviour. Nevertheless, drivers' reaction times increase significantly when using a mobile phone

[1] Corresponding Author: Eleonora Papadimitriou (e.papadimitriou@tudelft.nl)
[*] At the time that this research was carried out, this author was Research Associate at the National Technical University of Athens

(Consiglio et al., 2003), and little or no benefit of hands-free over handheld mobile phone has been validated (Caird et al., 2008; Törnros & Bolling, 2006). Overall, there is a statistically significant increase in crash risk when using a mobile phone while driving, namely almost three times the crash risk when a mobile phone is not used (Elvik, 2011; Backer-Grøndahl et al., 2011, Klauer et al., 2014).

The most common methodologies applied for the assessment of risks related to mobile phone use while driving are: (i) driving simulator experiments (Papantoniou et al., 2014), (ii) naturalistic driving experiments (e.g. Fitch et al., 2014; Hickman & Hanowski, 2012), and (iii) contributory factors analysis of actual crash records (Consiglio et al., 2003; Elvik, 2011; Backer-Grøndahl et al., 2011). Each method presents different advantages and limitations, however results are fairly consistent regardless of the study method. However, naturalistic driving studies tending to indicate higher benefits from hands-free mobile phone use (Ziakopoulos et al. 2018).

Naturalistic driving experiments provide a wide perspective of understanding normal microscopic travel and driving behaviour in normal conditions. A naturalistic study (Regan et al., 2012) can help to: a) estimate accident risk, b) study the interaction between road/ traffic conditions and driver's behaviour, c) understand the interaction between car drivers and vulnerable road users, d) specify the relationship between driving patterns and vehicle emissions or fuel consumption, and many other aspects of traffic participation.

There are numerous naturalistic driving studies dealing with mobile phone use and driver distraction. For instance, Fitch et al. (2015) analysed data from a naturalistic driving study that recorded 204 drivers using video cameras and vehicle sensors for an average of 31 days, and identified tasks associated with increased eyes-off-the-road behaviour. Klauer et al. (2014) used instrumented vehicles to analyse the crash or near-crash risk of novice drivers at the presence of secondary tasks, including the use of mobile phone. Dingus et al. (2016) used a naturalistic driving dataset comprising 905 injury and property damage crash events, and calculated a risk two times higher when mobile phone is used. Simmons et al. (2016) used 7 sets of naturalistic driver data to assess the effects of distracting behaviors in a random-effects meta-analysis including car drivers, non-commercial drivers of light vehicles, and commercial drivers of trucks and buses.

In the recent years, emerging methods for monitoring drivers on the basis of in-vehicle sensors are exploited for the collection of naturalistic data within everyday driving (Horrey et al. 2012; Vaezipour et al., 2015). The advantage of these sensors, compared to traditional "heavy" vehicle instrumentation of earlier naturalistic driving experiments, is the lighter and relatively low-cost equipment, combined with the new possibilities offered by information and communication technologies for data transmission and processing.

An initial relevant method of monitoring driving concerned OBD (On Board Diagnostics) recorders that are connected with the car engine. This was tested for a number of emerging telematics applications, including traffic management, accident detection and emergency response, monitoring of fuel consumption and emissions, monitoring of hybrid electric vehicles, monitoring of professional drivers etc. (Zaldivar J., et al. 2011; Yang et al., 2013). It is also used in innovative insurance schemes (UBI - User Based Insurance) on pricing on the basis of distance travelled or driving behaviour metrics provided by the OBD unit (Tselentis et al., 2017), as well as in other social gamification schemes of monitoring and 'rewarding' drivers for safe or environment-friendly driving (Musicant & Lotan, 2016; Munoz-Organero & Corcoba Magana, 2014).

More recently, smartphone sensors are tested as in-vehicle sources of driving behaviour metrics; indicatively, technology sensors that are integrated in contemporary smartphones are the accelerometer, the gyroscope, the magnetometer and the GPS. The data recorded by these sensors can be processed to yield meaningful driving metrics. These can be used to draw the profile of the driver (Araujo et al., 2012; Toledo et al., 2008; Johnson et al., 2011; Hong et al., 2014) and further to communicate this profile to the driver, in order to eventually improve road safety levels by increasing self-awareness and motivation (Toledo et al., 2008), often through the application of premiums based on driving behaviour (Tselentis et al., 2017).

Several existing studies have shown promising results as regards the analysis of certain risk factors i.e. speeding, aggressiveness etc. (Vlahogianni & Barmpounakis, 2017) through smartphone data collection and processing. However, to the best of the authors' knowledge, this is the first attempt to understand behaviours and risks related to the use of the mobile phone while driving, on the basis of data collected from smartphone sensors.

1.2. Objectives and Methods

The objective of this research is to explore driving behaviour during mobile phone use on the basis of detailed driving analytics collected by means of smartphone sensors. For that purpose, this research uses data collected from a naturalistic driving experiment with a sample of one hundred drivers. Using driving metrics calculated from the smartphone sensors data, a statistical analysis is carried out for correlating the use of mobile phone with other driving behaviour indicators, namely by means of mixed effects binary logistic regression. A general model was developed for all trips, as well as separate models for trips on different types of roads (urban, rural, highway).

2. Data collection

2.1. Overview

The data were collected through an innovative data collection scheme, developed by the OSeven Telematics Company (www.oseven.io), which records personalized driving behaviour analytics in real time, using smartphone sensors. An integrated system is used for the recording, collection, storage, evaluation and visualization of driving behaviour data, using smartphone applications and advanced Machine Learning (ML) algorithms. The system includes specially developed smartphone applications (apps) for data collection and transmission, as well as for providing feedback to the participants on their driving behaviour. The steps described below for data processing are exclusively performed by OSeven Telematics and do not constitute part of this study. More details on the data processing steps and cannot be provided since they are intellectual property of the company. However, the main features of the system are outlined below.

The data are stored in the OSeven backend system using advanced encryption and data security techniques, in compliance with the national laws and EU Directives for the protection of personal data. The APIs (Application Programming Interfaces) used support user authentication and encryption to prevent unauthorized data access. The data used in this research were derived from the OSeven database and provided by the company for the purposes of this study. The dataset concerns an anonymous sample of one hundred drivers in Athens, Greece, during a 4-months timeframe, including several thousands of trips.

2.2. Platform

A smartphone app is developed to record driver behaviour using the hardware sensors of the smartphone device, and a variety of APIs is exploited to read sensor data and temporarily store them to the smartphone's database before transmitting them to the central (backend) database.

The data recording is initiated automatically in the smartphone app when a driving status is recognized, and again it stops automatically when a non-driving status is recognized. The frequency of the data recording varies depending on the type of the sensor, with a minimum value of 1Hz. Trip recording also continues if the vehicle is idled for five minutes, to consider the case that the driver resumes a trip after a few minutes stop. All extra information collected after the end of driving is discarded. The basic operating frame of the data flow is shown in Figure 1.

Figure 1 to be inserted here

The recorded data come from various smartphone sensors and data fusion algorithms provided by Android (Google) and iOS (Apple). Indicatively, technology sensors integrated in the mobile phone are the Accelerometer*, the Gyroscope*, the Magnetometer and the GPS (speed, course, longitude, latitude). Fusion Data provided by iOS and Android include yaw, pitch, roll, linear acceleration* and gravity* (elements marked with an asterisk "*" sign refer to x, y, z components). After the end of each trip, the application is transmitting all data recorded to the central database of the OSeven backend office via an appropriate communication channel, such as a Wi-Fi network or cellular network (upon user's selection) e.g. 3G/4G (online options). The total volume of data transmitted for an average driver is estimated at around 50Mb / month.

The data collected are highly disaggregated in 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, 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 (see Figure 2). It is highlighted that all data are received from OSeven in an anonymized form.

Figure 2 to be inserted here

The main steps of the calculation procedure in the OSeven Platform are presented below:

i. Data filtering and outlier detection (all data that cannot be considered reliable are discarded)
ii. Data smoothening for the parameters that it is required (when abnormal outlier values are observed)
iii. Identification of speeding regions (duration of speeding, exceedance of speed limit - calculated on the basis of speed limit data from map providers e.g. Google, OSM, etc.)
iv. Identification of Harsh Acceleration/ Braking/ Cornering events
v. Identification of Mobile Usage (talking, texting, internet navigation) using only the smartphone sensors without any access to the user's activity
vi. Identification of Risky Hours Driving (distance travelled between 12am and 5am)
vii. Determination if the user is the driver or the passenger during the trip

viii. Calculation of the driving scores.

A variety of different metadata are eventually calculated, including 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)
- Weather conditions (under development, on the basis of integration with weather data providers)
- Trip purpose (set by the driver himself by using the smartphone app)

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

- Speeding (duration of speeding, speed limit exceedance etc.)
- Number and severity of harsh events
- Harsh braking (longitudinal acceleration)
- Harsh acceleration (longitudinal acceleration)
- Harsh cornering (angular speed, lateral acceleration, course)
- Driving aggressiveness (e.g. braking, acceleration)
- Distraction from mobile phone use (Mobile use is considered to be any smartphone usage by the driver e.g. talking, texting etc.).

These indicators along with other data (e.g. from map providers) can be subsequently exploited to calculate individual driver statistics, on all road networks (urban, rural, highway, etc.) and under various driving conditions, enabling the creation of a large database of individual trip / driver characteristics. Driving behaviour information and feedback can be potentially communicated back to drivers by means of a dedicated smartphone app as the one shown in Figure 3.

Figure 3 to be inserted here

3. Analysis method

The variable of interest in the present analysis is the use of mobile phone while driving. This was available either as a share of trip time during which mobile phone was used, or as a binary variable for the entire trip (yes / no). The latter case was selected for modelling in the present research.

Typically, a binary logistic regression estimates the probability that a characteristic is present (e.g. estimated probability of "success") given the values of explanatory variables; $\pi = Pr(y = 1|X = x)$. 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:

$$logit(\pi_i) = logit\frac{\pi_i}{1-\pi_i} = \beta_0 + \Sigma\beta_i x_i \quad (1a)$$

And the related outcome (event occurrence):

$$y_i = \beta_0 + \Sigma\beta_i x_i + e_{0i} \quad (1b)$$

In the present dataset, however, there are repeated measurements (trips) over the same units (drivers). Due to these repeated measurements, the observations are no longer independent, as required by the assumptions of logistic regression. 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. A mixed model (or random effects model, or multi-level model) is a standard technique to be used in this context, i.e. through a statistical model containing both fixed effects and random effects. The formulation of the mixed effects model, assuming a random intercept reflecting the repeated measurements (i) over drivers (j), is as follows:

$$y_{ij} = \beta_{0j} + \Sigma\beta_i x_{ij} + e_{0ij} \qquad (2a)$$

$$\beta_{0j} = \beta_0 + u_{0j} \qquad (2b)$$

It is noted that the intercept in the outcome equation (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 equation (2a) is assumed to follow a logistic distribution.

4. Results

4.1. Descriptive Analysis

For the purposes of this research, driving exposure and behaviour indicators on a trip basis were used, from a database of 18,850 trips from a sample of 100 drivers. For each trip, the share travelled on different road networks (urban, rural or highway) are also provided, and the vast majority of indicators are available separately for each type of road network within a trip.

The key indicator of interest for the purpose of this research is the use of mobile phone during the trip. Additional basic indicators (i.e. possibly suggesting risky or reckless behaviour) are: exceeding the speed limit (share of time / distance over the speed limit, share of speed over the speed limit), and harsh manoeuvres / events (including harsh accelerations, harsh braking, harsh cornering). On the basis of the literature review results, it is assumed that the use of mobile phone while driving is correlated with drivers' speeding behaviour and harsh events, and may thus be explained by changes in those variables.

Exploratory descriptive analysis of the data reveals some potentially interesting patterns. For example, the share of mobile phone use is found to be more frequent on urban roads (6.9 %), less so on rural roads (5.1 %) and very low on highways (0.9 %), which is not surprising as trips on highways are usually much longer ones. Figure 4 presents the total number of harsh events (including accelerations, braking and cornering) per trip distance, against the share of mobile phone use during the trip. Overall, the total number of harsh events per trip distance ranges around 0.4 and reaches a maximum value equal to 0.45 when the share of mobile phone use raises to 50-60% of the trip duration. It is noted that, the shorter the length of the trip, the higher the share of use of the mobile phone, possibly because longer journeys are more likely to be made on rural roads / highways. 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. A possible interpretation is that small duration mobile phone conversations may be less likely to result in harsh events. As conversation duration increases, it is more likely that the driver will make a harsh manoeuvre due to multi-tasking. For very long conversations, however, it is possible that a successful behavioural adaptation to multi-tasking eventually takes place.

Figure 4 to be inserted here

A nonlinear relationship appears to exist between the share of mobile phone use per trip and the speeding behaviour of drivers. Figure 5 shows that the share of mobile phone use per trip is somewhat higher on trips with average speed between 60 and 90 km/h (peaking between 80 and 90 km/h), and lower on trips with average speed lower than 60 km/h. This is an unexpected result since lower speeds are generally observed in urban areas, where mobile phone use is more frequent. It is possible, however, that mobile phone use in urban areas mostly takes place e.g. on urban arterials, where speed limits and speeds are higher, and less so on secondary / residential roads where the road environment is more complex (e.g. more junctions, more traffic signs, higher presence of pedestrians). In any case, this finding needs some further investigation.

The share of mobile phone use is much lower for average trip speed higher than 90 km/h, as is the case for main rural roads and highways.

Figure 5 to be inserted here

The exploratory analysis of the data suggested that the share of mobile phone use is indeed correlated with road characteristics, as well as with other driving behaviour metrics, namely speeding and harsh events. The extent to which mobile phone use (yes / no) can be associated with / predicted on the basis of such driving metrics, is investigated through the models developed.

4.2. Models development

The categorical dependent variable is the use or no use of mobile phone while driving, whereas the explanatory variables are the driving behaviour and exposure metrics described previously. The fixed effects are attributed to the explanatory variables and the random effect is attributed to the model intercept. A global model is developed for all road types, as well as separate models for each road type: urban, rural and highway.

The best performing model in each case was defined as follows: first, univariate effects were tested (i.e. each explanatory variable separately); for the statistically significant variables, correlation tests were performed, to avoid entering in the model any strongly correlated pair of explanatory variables. The selection of the optimal model took into account the statistical significance of variables and the overall fit of the model. The final models are presented in Table 1(a). 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 1 to be inserted here

The modelling results (Table 1) reveal the following as regards the correlation of driving exposure on mobile phone use:

- Driving during morning rush hours increases the odds of mobile phone use during the trip, possibly because drivers may make / receive more calls during morning rush hours (e.g. work related calls); the effect appears to be higher in urban areas and highways than in rural areas, as the respective parameters (B) are 0.233, 0.373 and 0.110, corresponding to odds ratios exp(B) of 1.26, 1.45, and 1.12. These values show the increase in log-odds of mobile phone use during the morning rush compared to the rest of the day. In other words, 1.26 is the odds ratio of the odds of mobile phone use during morning rush over the odds of mobile phone use during all other periods of the day except morning rush.
- Accordingly, driving during afternoon rush hours reduces the odds of mobile phone use during the trip. In contrast to morning rush, the effect of afternoon rush appears to be higher in rural areas than the other areas examined.
- The length of the trip (driving time) was not found to affect the odds of mobile phone use in the model for all road types. However, in all three separate models, a higher duration of the trip (urban roads and highways, both having odds ratio exp(B=-0.001) equal to 0.999 or a higher share of trip duration on the specific road type (rural roads with an odds ratio exp(B=-2.784) of 0.062) is associated with lower odds of mobile phone use during the trip.

Moreover, the following are found as regards the association of driving behaviour metrics with mobile phone use:

- Average speed per trip was found to be negatively associated with the odds of mobile phone use on all road types (odds ratio exp(B=-0.004) is equal to 0.996), confirming existing studies. The effect is significant only in the global model of all road types.
- The effect of exceeding the speed limit was found to be more sensitive than that of average speed. The average percentage exceedance of speed limits reduces the odds of mobile phone use; in general, drivers who are speeding more, are less likely to use their mobile phone during the trip, especially on highways (odds ratio exp(B=-2.512) equals 0.081). An exception is the urban road environment, where a relatively low positive relationship is found between higher exceedance of speed limit and the odds of mobile phone use.
- The higher the number of harsh events per trip distance, the lower the odds of mobile phone use (parameter B is negative for all road types except for highways); the literature suggests that drivers reduce speed while distracted, and therefore are less prone to harsh events. Again, there is an exception, namely while driving on highways, where a higher number of harsh events is associated with higher odds of mobile phone use. This may be interpreted by the higher speeds on highways, which may not be easily compensated.
- The variable average angular speed (measured in ^{o}/s) reflects the smoothness of cornering manoeuvres during the trip. This particular type of harsh event was found statistically significant in the global model, suggesting that the higher the angular speed, the lower the odds of mobile phone use (odds ratio exp(B=-0.058) equals to 0.94).

Table 1(b) presents the classification of outcomes as per the final model for all road types. It can be seen that more than 70% of actual cases where mobile phone was used during a trip are correctly identified by the model. “False positives”, i.e. cases falsely classified as mobile phone used are 28.6%. However, classifications are less successful in the separate models; the true cases of mobile phone use are correctly predicted by 58%, 43% and 29% on urban roads, on rural roads and on highways respectively. It is noted that the “false positives” are minor shares of the classified cases in all three separate models, suggesting that the driving metrics found to be statistically significant may accurately predict cases of “not using mobile phone”, but not so accurately “using mobile phone” cases.

5. Discussion

The results from the interpretation of the estimated parameters of the models can be summarized as follows: Exposure metrics significantly associated with the odds of mobile phone use while driving are the trip length and the driving off-morning rush. Exceeding speed limits, the total number of harsh events, and the harsh cornering in particular, are all associated with the odds of someone using the mobile phone, with a general patterns of less speeding and smoother driving being indicative of mobile phone use.

These results refer to the use of mobile phone as a binary variable (yes/no) at the trip level. It is interesting to note, however, that descriptive analysis using the share of mobile phone use over the trip duration, revealed that longer conversations may be related to an increase in harsh events. Future statistical analysis will also consider the share of mobile phone use as dependent variable, for further insights into distracted driving behaviour.

Overall, it appears that mobile phone use can be “detected” in the absence of other risky or reckless behaviours, and not at their presence, which is in line with the findings of previous studies. The detection of the use of mobile phone while driving on the basis of other driving metrics can be made quite accurately when all trips are examined together, as well as for trips in urban areas. There are some specific patterns identified that differentiate between urban, rural and highway driving while using mobile phone, but the overall model is considered more reliable than the road type specific models.

The current lack of data on driver characteristics (e.g. age, gender, etc.) certainly limits the potential of further explanation and more correct detection of distracted driving, as existing studies suggest that these variables would be important additional predictors. Nevertheless, the proposed general model may classify correctly a fair share of distracted driving cases, even though these “predictions” are based solely on driving exposure and behaviour metrics and no individual driver characteristics. It is important to highlight that there is little or no previous experience on analysing and predicting mobile phone use through microscopic driving behaviour metrics collected from such a portable in-vehicle device, and therefore the results of the present research cannot be directly compared to those of existing literature.

The present analysis has some other limitations as well. The use of relatively aggregate data (e.g. average metrics per trip) may have compromised the identification of more detailed patterns corresponding to distracted driving. The more disaggregate data available after the cleaning and first processing of the raw data may allow for more detailed analysis; however, in this case completely different Big Data analysis methods are required, mostly non parametric methods. The present paper applied a relatively “traditional” modelling technique and the results are promising that more disaggregate analysis would provide further insights.

Moreover, the analysis focuses on mobile phone conversation in handheld mode, as no data could be made available at the time of the research on hands-free conversation, texting, browsing etc., but this will be pursued in the near future.

As regards the data collection system, a number of observations can be made: the use of smartphone sensors alone has some limitations compared to full vehicle instrumentation in naturalistic driving experiments, most importantly the lack of data on vehicle headways, driver reaction time (e.g. brake response) or video data (e.g. eye-glances behaviour). The main purpose of the models developed is the analysis and potential detection of mobile phone use through other driving metrics. In this context, the metrics collected appear to provide insights on a number of additional and more detailed driving behaviour factors associated with mobile phone use, making this "lighter" approach quite promising. It is possible that rapid developments in smartphone technologies (e.g. dual cameras) may boost the possibilities for additional driving behaviour metrics and additional analyses.

6. Conclusions

This paper aimed to investigate the potential of analysing driving offences on the basis of driver exposure and behaviour metrics collected by smartphone sensors, focusing on the case of mobile phone use while driving. The question is particularly important as mobile phone use while driving is known as a major and persistent risk factor (alongside speeding, alcohol, etc.). Given that the penetration and use of mobile phones is expected to further increase, together with their numerous emerging functionalities and apps, driving risks may also increase.

The main implication of the findings are the following: First, the results reveal correlations of mobile phone use while driving with specific driving behaviour and exposure metrics, at a more detailed level than existing studies (e.g. exceeding speed limit, harsh cornering). Second, the findings suggest that, by means of driving behaviour metrics, one may potentially "detect" phone use. The driving metrics of interest can be collected by means of a variety of sensors: the smartphone itself, an OBD device etc. Therefore, further research should focus on the investigation and validation of this potential. Eventually, real-time detection of mobile phone use may support the implementation of innovative enforcement and awareness raising schemes, through driver notification and alert (either in real-time or post-trip) aiming to reduce the use of mobile phone while driving.

At a more general level, the expected increase in the use of modern smartphones, and the rapid digitalization of so many every day activities, may be seen as an opportunity to exploit the wealth of data that can be made available by smartphone sensors, and the new possibilities for data transmission and processing, in order to identify new ways of tackling and mitigating the prominent risk factors and improve road safety.

For instance, by identifying the key risk factors while driving, vehicle manufacturers may be able to develop new systems that will directly improve the safety of vehicle occupants and other road users through primary and secondary safety features. One specific application area for the industry relates to the development of targeted advanced automatic driver distraction recognition and prevention systems. Moreover, vehicle insurance companies may exploit smartphone data within their UBI schemes and reward cautious drivers with reduced premiums. Telematics and IT groups may develop new smartphone apps and social gamification schemes for road safety targeted directly to the individual drivers.

The present research contributes a preliminary example of explaining and detecting mobile phone use while driving on the basis of a relatively small number of driver exposure and behaviour metrics. Despite some limitations, the results suggest that such analysis is feasible, and further research should focus on the improvement of the accuracy of the models, by exploring more variables and alternative modelling techniques. An intuitive next step of this research is the validation of the models with a new dataset and a simulation-based estimation of model predictions, and this will be pursued in the near future.

Acknowledgments

The authors would like to thank OSeven Telematics (www.oseven.io), for providing all necessary data to accomplish this study.

References

Araujo, R., Igreja Â., de Castro R., Araújo R.E. (2012). Driving Coach: A Smartphone Application to Evaluate Driving Efficient Patterns. IEEE Intelligent Vehicles Symposium.

Backer-Grøndahl A., Sagberg F. (2011) Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science 49(2), 324-330.

Caird, J. K., Willness, C. R., Steel, P. & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. Accident Analysis & Prevention 40, 1282-1293.

Consiglio, W., Driscoll, P., Witte, M. & Berg, W.P. (2003). Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. Accident Analysis & Prevention 35, 495-500.

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, 201513271.

Elvik, R. (2011). Effects of mobile phone use on accident risk: Problems of meta-analysis when studies are few and bad. Transportation Research Record No 2236, 20-26.

Fitch, G., Grove, K., Hanowski, R., & Perez, M. (2014). Compensatory behavior of drivers when conversing on a cell phone: Investigation with naturalistic driving data. Transportation Research Record: Journal of the Transportation Research Board, (2434), 1-8.

Fitch, G. M., Bartholomew, P. R., Hanowski, R. J. & Perez, M. A. 2015. Drivers' visual behavior when using handheld and hands-free cell phones. Journal of safety research, 54, 105.

Haque M.M., Washington S. (2015). The impact of mobile phone distraction on the braking behaviour of young drivers: A hazard-based duration model. Transportation Research Part C: Emerging Technologies 50, 13-27.

Hickman, J. S., & Hanowski, R. J. (2012). An assessment of commercial motor vehicle driver distraction using naturalistic driving data. Traffic injury prevention, 13(6), 612-619.

Hong J-H, Margines B., Dey A.K. (2014). A Smartphone-based Sensing Platform to Model Aggressive Driving Behaviors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 4047-4056, Toronto, Canada.

Horrey, W. J. & Wickens, C. D. (2006) Examining the impact of cell phone conversations on driving using meta-analytic techniques. Human Factors 48, 196-205.

Horrey W.J., Lesch M.F., Dainoff M.J., Robertson M.M. & Noy Y.I. (2012). On-board safety monitoring systems for driving: Review, knowledge gaps, and framework. Journal of Safety Research 43;49-58.

Johnson D.A, Trivedi M.M. (2011). Driving Style Recognition Using a Smartphone as a Sensor Platform. 14th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. New England journal of medicine, 370(1), 54-59.

Munoz-Organero M, & Corcoba Magana V. (2014). The impact of using gamification on the eco-driving learning. Ambient Intelligence- Software and Applications, Advances in Intelligent Systems and Computing 291; 45-52.

Musicant O., & Lotan T. (2016). Can novice drivers be motivated to use a smartphone based app that monitors their behavior? Transportation Research Part F – Psychology & Behaviour 42(3); 544-557.

Papantoniou P., Papadimitriou E., Yannis G. (2014). Assessment of Driving Simulator Studies on Driver Distraction. Advances in Transportation Studies 35, 129-144.

Regan, M. A., Williamson, A., Grzebieta, R., & Tao, L. (2012, August). Naturalistic driving studies: literature review and planning for the Australian naturalistic driving study. In Australasian college of road safety conference 2012, Sydney, New South Wales, Australia.

Simmons, S. M., Hicks, A. & Caird, J. K. (2016). Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis. Accident Analysis & Prevention 87, 161-169.

Toledo, T., Musicant O., Lotan T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behaviour. Transportation Research Part C: Emerging Technologies 16(3), 320-331.

Törnros J., Bolling A. (2006). Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. Transportation Research Part F: Traffic Psychology and Behaviour 9(4), 298-306.

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.

Vaezipour A., Rakotonirainy A., & Haworth N. (2015). Reviewing in-vehicle systems to improve fuel efficiency and road safety. Procedia Manufacturing 3; 3192-3199.

Vlahogianni, E. I., & Barmpounakis, E. N. (2017). Driving analytics using smartphones: Algorithms, comparisons and challenges. Transportation Research Part C: Emerging Technologies, 79, 196-206.

Yang Y., Chen B., Su L., Qin D. (2013). Research and Development of Hybrid Electric Vehicles CAN-Bus Data Monitor and Diagnostic System through OBD-II and Android-Based Smartphones. Advances in Mechanical Engineering, Article ID 741240.

Zaldivar J., et al. (2011). Providing Accident Detection in Vehicular Networks through OBD-II Devices and Android-based Smartphones. 5th IEEE Workshop On User MObility and VEhicular Networks.

Ziakopoulos, A., Theofilatos, A., Papadimitriou, E., Yannis, G. (2018), Distraction - Cell Phones - Hands Free, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafetydss.eu on 10 02 2019

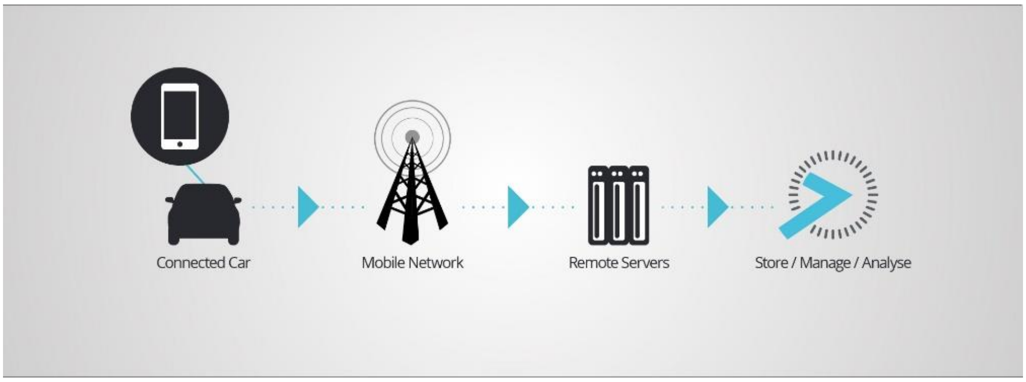

Figure 1. The OSeven data flow system

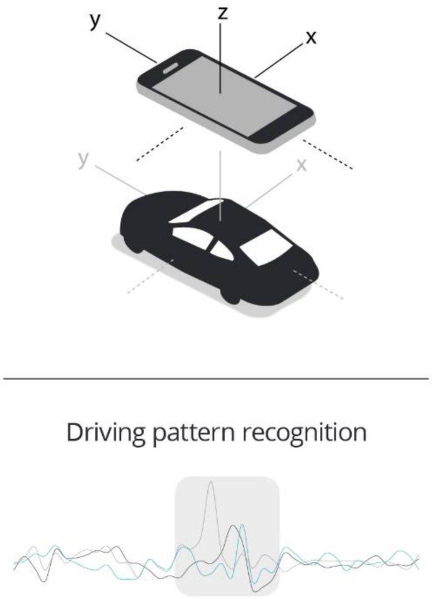

Figure 2. Driving pattern recognition by means of Machine Learning algorithms on Yaw, Pitch & Roll data

Figure 3. Driving behaviour indicators and OSeven driver feedback app

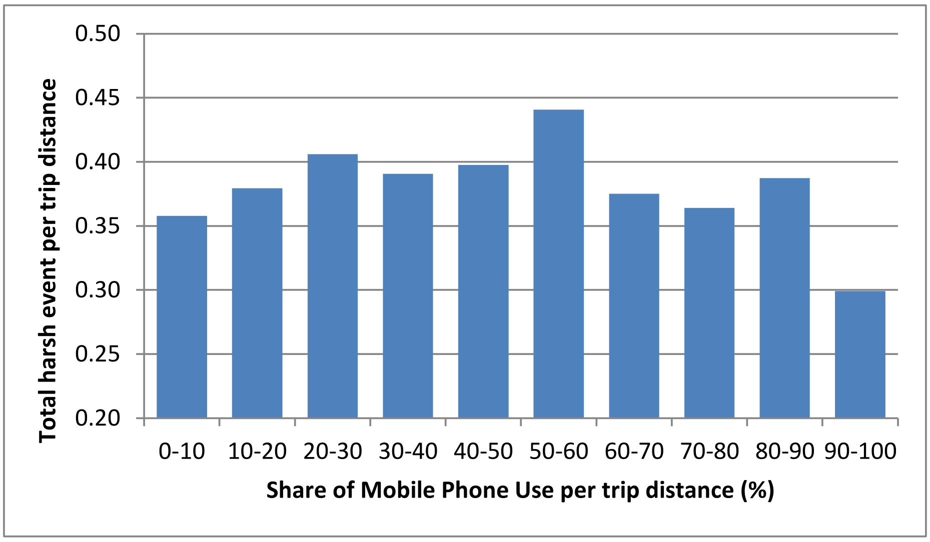

Figure 4. Total number of harsh events per trip distance in relation to the share of mobile phone use during the trip

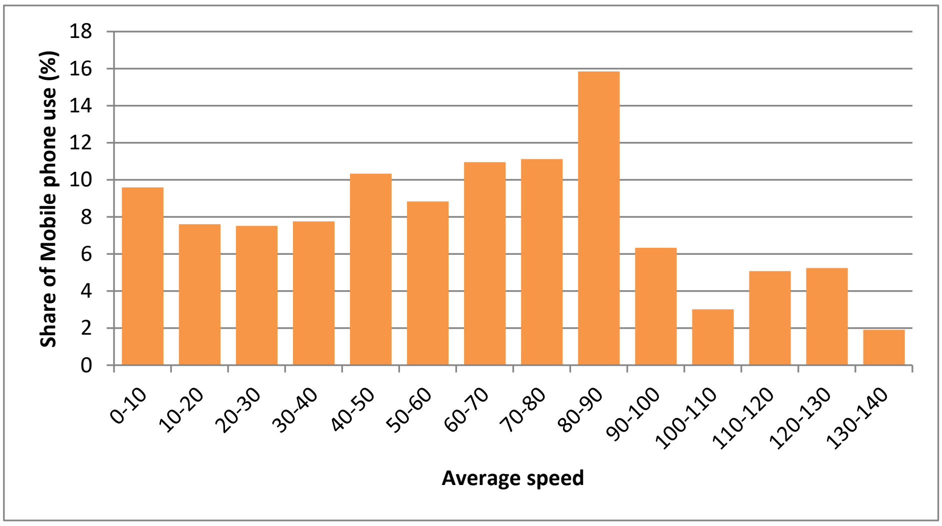

Figure 5. Share of mobile phone use per trip in relation to average speed

Table 1. (a) Parameter estimates of the mixed binary logistic models for all road types, and separately for urban roads, rural roads and highways - (b) Outcomes Classification Table for all road types

(a)

Parameter estimates	**All road types**		**Urban roads**		**Rural roads**		**Highways**	
Fixed effects	**B**	**P-value**	**B**	**P-value**	**B**	**P-value**	**B**	**P-value**
Constant	1.094	<0.001	1.488	<0.001	2.443	<0.001	4.138	<0.001
Morning rush	0.130	0.006	0.233	<0.001	0.110	0.044	0.373	0.002
Afternoon rush	-0.262	<0.001	-0.111	0.008	-0.206	<0.001	-0.161	<0.001
Average percentage of speed over the speed limit	-0.334	0.027	0.395	<0.001	-0.938	<0.001	-2.512	<0.001
Average speed	-0.004	0.001	-	-	-	-	-	-
Average angular speed	-0.058	<0.001	-	-	-	-	-	-
Total Harsh events	-0.064	<0.001	-0.034	<0.001	-0.084	<0.001	0.148	<0.001
Time driving	-	-	-0.001	<0.001	-	-	-0.001	-
Percentage of time driving in road type	-	-	-	-	-2.784	<0.001	-	-
Random effect (variance of random intercept)	1.261	<0.001	1.346	<0.001	1.153	<0.001	1.429	<0.001
Number of observations (trips)	18,853		18,853		18,853		18,853	
Number of drivers	100		100		100		100	
AIC*	86,301		88,375		93,315		142,526	

*Akaike Information Criterion

(b)

Mobile Phone use	**Predicted (%)**	
Observed	No	Yes
No	71.4	28.6
Yes	29.9	70.1