Monitoring and Improving Driving Behaviour of Motorcyclists Through an Innovative Smartphone Application

Armira Kontaxi, National Technical University of Athens, Greece, *akontaxi@mail.ntua.gr*, **Apostolos Ziakopoulos,** National Technical University of Athens, Greece, **Panagiotis Papantoniou,** National Technical University of Athens, Greece, **George Yannis**, National Technical University of Athens, Greece, **George Kostoulas**, OSeven Single Member Private Company, Greece

ABSTRACT

The objective of the present study is twofold: (i) to explore driving behaviour of motorcyclists while speeding, based on detailed driving analytics collected by smartphone sensors, and (ii) to investigate whether personalized feedback can improve motorcyclist behaviour. The objectives are achieved through a naturalistic driving experiment with a sample of 20 motorcyclists based on a smartphone application developed within the framework of the BeSmart project. Using risk exposure and driving behaviour indicators calculated from smartphone sensor data, Generalized Linear Mixed-Effects Models are calibrated to correlate the percentage of driving time over the speed limit with other driving behaviour indicators. Results indicate that the parameters of trip duration, distance driven during risky hours, morning peak hours and the number of harsh accelerations have all been determined as statistically significant and positively correlated with the percentage of speeding time. Additionally, driver feedback and afternoon peak hours are statistically significant and negatively correlated with the percentage of speeding.

Keywords: road safety, motorcyclists, driver monitoring, naturalistic experiment, smartphone application, Generalized Linear Mixed-Effects Models

1 INTRODUCTION

1.1 Background

Motorcyclists constitute a vulnerable road user group with up to 30 times higher fatality rates compared to passenger cars (Johnston et al., 2008). In 2017, motorcyclists accounted for 18% of the total number of road deaths in the EU countries; specifically, about 3,850 riders (drivers and passengers) of motorcycles and about 600 riders of mopeds were killed in EU countries in traffic crashes (CARE, 2020). During 2016, Greece had the highest rate of motorcycle fatalities per million population (24) and per total road crash fatalities (32) in EU-28 Countries (European Commission, 2018).

Specific factors affecting the crash injury severity of motorcyclists have been determined in the literature as well: Albalate & Fernandez-Villadangos (2010) identified gender, excess speed, road width, and alcohol consumption as factors affecting powered two-wheeler (PTW) injury severity. Theofilatos & Ziakopoulos (2018) determined that traffic and speed variations increase PTW injury severity, while increased truck proportions in the traffic mix were found to relatively reduce injury severity, possibly due to behavioural adaptations on behalf of PTW riders.

Behavioural issues are major moderating factors to both the frequency and the severity of motorcycle crashes. Speeding, sensation seeking, aggression, perceived risk, errors, violations and attitudes towards road safety are considered to be crucial behavioural critical risk factors (Vlahogianni et al., 2012; Theofilatos and Yannis, 2015). Therefore, the accurate monitoring of motorcyclist driving behaviour is of high importance. Although motorcycle crashes have been widely investigated by researchers, the lack of detailed naturalistic driving data remains a disturbing obstacle for the scientific community. Several existing studies have shown promising results as regards the analysis of motorcyclist driving behaviour by means of naturalistic experiments (Espié et al., 2013; Vlahogianni et al., 2014) or even video footage from unmanned aerial vehicles (UAV) (Barmpounakis et al., 2017). However, to the best of the authors' knowledge, this is the first attempt to understand behaviours and risks related to motorcyclists on the basis of data collected from smartphone sensors.

1.2 Objectives

In light of the aforementioned, the objective of the present study is twofold: (i) to explore driving behaviour of motorcyclists while speeding, based on detailed driving analytics collected by smartphone sensors, and (ii) to investigate whether personalized feedback can improve driver behaviour. For that purpose, high-resolution smartphone data collected from a naturalistic driving experiment with a sample of 20 motorcyclists are utilised. Using risk exposure and driving behaviour indicators calculated from the smartphone sensors data, a statistical analysis is carried out for correlating the percentage of driving time over the speed limit with other driving behaviour indicators, namely by means of Generalized Linear Mixed-Effects Models. For reasons of brevity, from now on the percentage of driving time over the speed limit will be mentioned as speeding percentage henceforth in the text.

2 Methodology

2.1 Experimental Design

In the framework of BeSmart Project, a naturalistic experiment spanning 12 months is conducted with different driver types participating, including the vulnerable road user group of motorcyclists. More specifically, the designed experiment consists of six different phases differing in the type of feedback provided to drivers. In the present paper, the authors focus on the two first phases of the experiment. The first phase lasted for 12 weeks; participants were asked to drive in the way they usually did, without receiving any feedback on their driving behaviour from the smartphone application. The purpose of this stage was to learn drivers' naturalistic driving characteristics, which provide a baseline for comparison. The second phase lasted for 10 weeks; participants were provided with personalized feedback, a trip list and a scorecard regarding their driving behaviour, allowing them to identify their critical deficits or unsafe behaviours.

2.2 The BeSmart Application

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. OSeven Telematics has also developed the solid integration platform for collecting, transferring raw data and recognizing the driving behaviour metrics via ML algorithms. The standard procedure that is followed every time a new trip is recorded by the application is clearly presented in Figure 1.

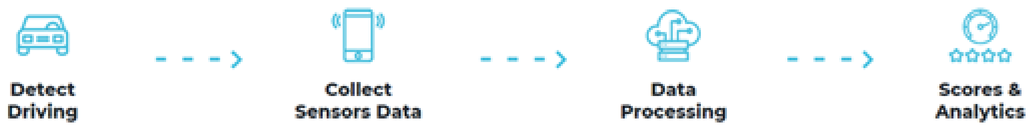

Figure 1 - The OSeven data flow system.

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. All this is done using state-of-the-art technologies and procedures in compliance with standing Greek and European personal data protection legislation (GDPR).

The available exposure indicators 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) and time of the day driving (peak hours, risky hours). Moreover, the driving indicators which can reliably quantify the risk associated with a specific driving behaviour consist of 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, driving aggressiveness (e.g. braking, acceleration), and eco-driving (smooth use of the accelerator, steering, transmission and brakes).

3 Methodology

The variable of interest in the present analysis is the fraction of speeding while driving. This quantity was available either as a share of trip time during which the speed limit was exceeded, or as a binary variable for the entire trip (yes/no). The first approach was selected for modelling in the present research. After transforming the speeding percentage per trip to an integer, Generalized Linear Models (GLMs) were implemented with a Poisson data distribution. GLMs are known to be better used when dealing with frequency (count) data (Lord & Mannering, 2010).

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

$$y_i \sim Poisson\ (\lambda_i) \quad (1)$$

And the linear predictor is:

$$log(\lambda_i) = \beta_0 + \beta_n x_n + \varepsilon \quad (2)$$

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

However, one may also consider that in the present dataset there are repeated measurements (trips) over the same units (drivers). Therefore, in order to capture personal driver traits, such as personality and experience, which affects their driving style, and thus the speeding percentage they exhibit, random effects are introduced to GLMs in order to extend them as Generalized Linear Mixed-Effects Models (GLMMs). Random effects in GLMMs are expressed as random variable coefficients (random slopes) or random intercepts. For a GLMM containing a random intercept and a random slope for a single independent variable j of the total n, Equation (2) would be formulated as:

$$log(\lambda_i) = \beta_{0i} + \beta_{ji}x_{ji} + \beta_{n-1}x_{n-1} + \varepsilon \tag{3}$$

Where β_{0i} and β_{ji} follow normal distributions centred at the value of their fixed counterparts:

$$\beta_{0i} \sim N(\beta_0, \sigma_{s,0}^2) \tag{4}$$

$$\beta_{ji} \sim N(\beta_j, \sigma_{s,j}^2) \tag{5}$$

4 RESULTS

Overall, during the first two phases of the experiment 3,853 trips from a sample of 20 motorcyclists have been recorded. However, for the present analysis it was decided that the final sample should consist only of drivers who have participated equally in both phases on terms of trips. As a result, from the 20 motorcyclists, 13 were ultimately selected creating a large dataset of 3,537 trips. Figure 2 illustrates a boxplot of the speeding percentage per experiment phase, allowing for an initial investigation of the response variable; there is a considerable presence of outliers of speeding time.

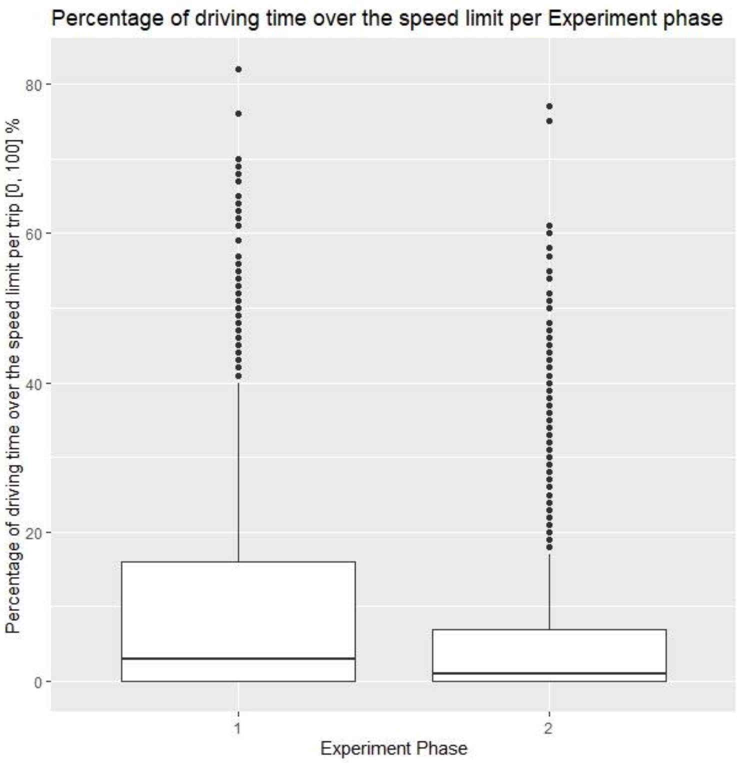

Figure 2 - Boxplots of the percentage of driving time over the speed limit per experiment phase

GLMMs were fitted in R-studio (with the lme4 package) via maximum likelihood and using z-factor scaling, following Bates et al. (2013). A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. The selected variables were chosen after taking into account the following: lowest AICc, high statistical significance of variables, low multicollinearity, and final rational interpretation of their impact on the dependent variable. After conducting log-likelihood test ANOVA comparisons, the most informative configuration of random effects was included both random intercepts and random slopes in the GLMMs to capture unique driver traits. Results of mixed effect selection are shown on Table 1; fixed effect results appear on Table 2.

Table 1 - Log-likelihood comparison of mixed effect selection

Model Family	Model Configuration	D.f.	LogLikelihood	χ^2	P(>χ^2)	Sig.
GLM	Fixed effects only [baseline]	7	-26227	–	–	–
GLMM	Fixed effects & Random Intercepts	8	-19390	13674.6	<2e-16	***
GLMM	Fixed effects, Random Intercepts & Random Slopes	10	-18547	1684.9	<2e-16	***

Significance codes: ‘***’: 0.000 | ‘**’: 0.001 | ‘*’: 0.01 | ‘.’: 0.05 | ‘ ’: ≥ 0.1

Table 2 - GLMMs for the percentage of driving time above the speed limit

Trip Parameter	Estimate	s.e.	p-value	Sig.	Rel. Risk Ratio
Intercept	1.898	0.276	<0.001	***	6.672
Exp. phase 2 (0=no feedback, 1=feedback)	-0.145	0.013	<0.001	***	0.865
Total trip duration (s)	0.194	0.095	0.042	*	1.214
Number of harsh accel. per trip (count)	0.248	0.005	<0.001	***	1.281
Trip distance during risky hours [22:00-05:00]	0.018	0.003	<0.001	***	1.018
Morning peak hour [06:00-10:00] (0=yes, 1=no)	0.067	0.015	<0.001	***	1.069
Afternoon peak hour [16:00-20:00] (0=yes, 1=no)	-0.286	0.015	<0.001	***	0.751

Significance codes: ‘***’: 0.000 | ‘**’: 0.001 | ‘*’: 0.01 | ‘.’: 0.05 | ‘ ’: ≥ 0.1

Modelling results regarding speeding percentage reveal some interesting findings; the parameters of trip duration, the distance driven during risky hours, morning peak hours and the number of harsh accelerations have all been determined as statistically significant and positively correlated with the percentage of speeding. In the same context, driving during the second phase of the experiment, as well as afternoon peak hours are statistically significant and negatively correlated with speeding percentage.

More specifically, the exposure metrics of trip duration, trip distance during risky hours and morning peak hours seem to increase speeding percentage by a factor of 1.214, 1.018 and 1.069 respectively. In other words, motorcyclists seem prone to speeding while driving under circumstances that increase their impatience and/or stress such as long trip durations, driving during hours of increased traffic conflicts, lane-splitting, hurrying while commuting, etc. Additionally, the driving behavioural parameter of harsh accelerations increases speeding percentage by a factor of 1.281, indicating the pattern of a stressful driving style.

Regarding the different experiment phases, providing motorcyclists with feedback about their driving performance leads to a remarkable decrease of speeding percentage by 14.5%. As explained above, during Phase 2 drivers received personalized feedback regarding their weak points, namely speeding and aggressive driving (harsh accelerations and harsh breakings) by means of a scorecard through the smartphone application. Therefore, the quantification of the positive effect of driver feedback on driving performance indicates new ways of improving road safety.

Furthermore, the visual representations of values of random intercepts and random slopes for total trip duration per driver are shown in Figure 3. Personal differences per driver from the fixed effect intercept and slope are thus included in the linear predictor.

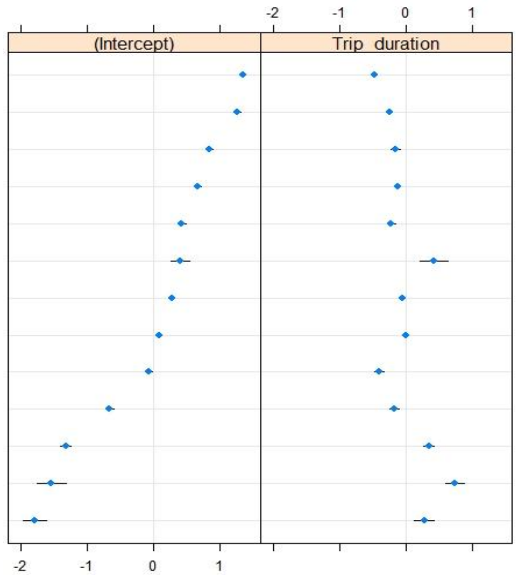

Figure 3 - Random Intercepts and Random Slopes for total trip duration

5 IMPACT

The outcomes of this study entail both scientific and social impacts. The present research contributes a preliminary example of the quantitative documentation of the impact of personalized driver feedback on one of the most important human risk factors; speeding. The ultimate objective when providing feedback to drivers is to: (i) trigger their learning and self-assessment process, thus enabling them to gradually improve their performance and (ii) monitor their evolution. The present results capture and quantify these positive effects of driver feedback, thus providing needed impetus for larger-scale applications as well as relevant policy interventions. State-of-the-art interventions can include approaches for driver training and support through innovative driver behaviour monitoring and feedback tools for different types of drivers, including the vulnerable road user group of motorcyclists. Regarding further research, microscopic data analysis of the collected database could be implemented through machine learning techniques and structural equation models. Future research will also focus on the different types of personalized feedback that will be communicated to motorcyclists in the next phases of the experiment, namely incentives within a social gamification scheme, with personalized target setting, benchmarking and comparison with peers.

ACKNOWLEDGEMENTS

This research is co-financed by the European Union - European Regional Development Fund (ERDF) and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) of the National Strategic Reference Framework (NSRF) - Research Funding Program: BeSmart - Multi-modal driver behaviour and safety support system on the basis of smartphone applications.

REFERENCES

Albalate, D., & Fernández-Villadangos, L. (2010). Motorcycle injury severity in Barcelona: the role of vehicle type and congestion. Traffic injury prevention, 11(6), 623-631.

Barmpounakis, E. N., Vlahogianni, E. I., & Golias, J. C. (2017). Identifying predictable patterns in the unconventional overtaking decisions of PTW for Cooperative ITS. *IEEE Transactions on Intelligent Vehicles*, *3*(1), 102-111.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7), 1-23.

Espié, S., Boubezoul, A., Aupetit, S., & Bouaziz, S. (2013). Data collection and processing tools for naturalistic study of powered two-wheelers users' behaviours. *Accident Analysis & Prevention*, *58*, 330-339.

European Commission, (2018). Traffic Safety Basic Facts on Motorcycles & Mopeds, European Commission, Directorate General for Transport, June 2018

Johnson, P., Brooks, C., & Savage, H. (2008). Fatal and serious road crashes involving motorcyclists (No. Monograph 20).

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation research part A: policy and practice, 44(5), 291-305.

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.

Theofilatos, A., & Yannis, G. (2015). A review of powered-two-wheeler behaviour and safety. International journal of injury control and safety promotion, 22(4), 284-307.

Theofilatos, A., & Ziakopoulos, A. (2018). Examining Injury Severity of Moped and Motorcycle Occupants with Real-Time Traffic and Weather Data. Journal of Transportation Engineering, Part A: Systems, 144(11), 04018066.

Vlahogianni, E. I., Yannis, G., & Golias, J. C. (2014). Detecting powered-two-wheeler incidents from high resolution naturalistic data. Transportation research part F: traffic psychology and behaviour, 22, 86-95.

Vlahogianni, E. I., Yannis, G., & Golias, J. C. (2012). Overview of critical risk factors in Power-Two-Wheeler safety. Accident Analysis & Prevention, 49, 12-22.