Rethinking transport
27–30 April 2020

Proceedings of 8th Transport Research Arena TRA 2020, April 27-30, 2020, Helsinki, Finland

Advanced driver monitoring using smartphone applications: The BeSmart project

Apostolos Ziakopoulos[a]*, Armira Kontaxi[a], George Yannis[a],
Petros Fortsakis[b], Kleanthis - Nikolaos Kontonasios[b], George Kostoulas[b]

[a]*National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str., GR-15773, Athens, Greece*
[b]*OSeven Telematics Limited, 27B Chaimanta Str., GR-15234, Athens, Greece*

Abstract

Driver monitoring involves the observation and recording of crucial driving tasks, such as speeding or distracted driver behavior. The BeSmart project aims to develop an innovative driver monitoring smartphone application and its calibration to real-world performance demands. A 200-driver naturalistic driving experiment spanning 18 months was initiated during May 2019, with different driver types participating. Several parameters are recorded from driver trips, such as speed and position. Processing of the collected big data leads to the pinpointing of less safe driver behavior, such as harsh events, speeding and distraction via smartphone interaction. An initial analysis of car driver trips during the familiarization phase of BeSmart was conducted. 11211 trips from 132 drivers were examined and clustered in three types: Aggressive, Speeding and Average. Generalized Linear Mixed-Effects Models were fitted to the trips of 76 car drivers who made frequent trips in order to model the frequencies of harsh events.

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

* Corresponding author. Tel.: +30-210-772-1575;
E-mail address: apziak@central.ntua.gr

1. Introduction

In light of rapid technological advancement during the previous decade, driver monitoring and its potential improvements remain a critical transport challenge. Driver monitoring involves the observation and recording of risky driving performance, such as speeding or driver behavior under distraction factors. Speeding is well-documented as a factor influencing crash occurrence and probability (WHO 2018; Elvik et al., 2019). Driver distraction is well documented in the literature, with several distraction sources having been extensively investigated such as distraction from various forms of mobile phone use, interaction with passengers, factors outside vehicles and other forms (e.g. Regan et al., 2011; Guo et al., 2017; Theofilatos et al., 2018).

Some of the recent developments in road safety are oriented towards driver monitoring through naturalistic driving. Examined approaches of that area include the employment of high-end technological solutions, such as motion and trajectory analysis from video images, exploitation of On-Board Diagnostics (OBD) and smartphone data collection (for instance Zaldivar et al., 2011; Nikias et al., 2012; Ma et al., 2018). The last method has several substantiated benefits such as allowing uninterrupted and swift data collection and wide application capabilities, with diminishing costs per examined driver (though up-front development costs are non-negligible).

In this environment, the aim of the present study is to showcase the conceptual framework and preliminary results of the BeSmart project. The objective of the BeSmart project is to develop an innovative and seamless Internet of Things application with tools to evaluate and improve the behavior and safety of all drivers (car drivers, powered two-wheelers, cyclists, professional drivers) along multi-modal trips. A wide naturalistic experiment is set to provide a multitude of parameters from driver monitoring in the form of big data, which will be analyzed to evaluate driver performance. The project is expected to provide insights and exploitation potential for wide acceptance by the end users of the proposed tools (individual drivers, scientific community, businesses and professionals, road management authorities, etc.).

In order to showcase the potential of the BeSmart project and its approach, an analysis of a sample of the initial data collected from phase 1 is conducted in this paper. More specifically, trips from 132 car drivers are analyzed collectively for the extraction of descriptive statistics and k-means cluster analysis. Furthermore, Generalized Linear Mixed-Effects Models (GLMMs) with the Poisson function are estimated using high-level trip data of frequent drivers, in order to estimate the frequency (counts) of harsh-acceleration and harsh-braking events.

2. Methodology

In order to achieve the research aim, which is the development of the driver monitoring application and its calibration to real-world performance demands, a testing ground was required. Thus, within the framework of the BeSmart project, a 200-driver naturalistic driving experiment is actively being planned at present. The experiment is scheduled for a duration of 18 months and was initiated during May 2019. Throughout its duration, participant drivers will install the application on their smartphone devices and drive as they would normally. The objectives of the experiment include, primarily, unobtrusive data collection for several parameters via smartphone sensors. These parameters include geographical position coordinates for vehicle location (latitude, longitude, altitude), speed, acceleration, mobile phone use, as well as speeding duration (referring to local road speed limits). Another family of parameters, crucial for understanding personalised driver behavior, is harsh events: number and severity of harsh events (braking and acceleration) is registered when accelerometers record values exceeding pre-determined thresholds. Algorithms of ecological driving detection are also included in the application. It is important to note that no other equipment or vehicle instrumentation whatsoever is going to be installed in the vehicles of the participants, providing overall a low-cost instrumentation, flexible approach.

Processing of the collected big data leads to the pinpointing of unsafe driver behavior, such as harsh events (including harsh braking or acceleration instances) as well as their intensity, harsh cornerings, driving aggressiveness (e.g. braking, acceleration), speeding and distraction via smartphone interaction. Overall driver behavior is then assessed to measure the overall safety performance of each driver, complementing the first pillar of the application which regards driver monitoring. BeSmart is designed to analyze participants from multiple transport modes, including car drivers, professional drivers, powered-two-wheeler riders and bicycle riders. Furthermore, specific care was taken during the recruitment phase to ensure a representative sample through participation of drivers of varying age and gender in the experiment. Despite the very high resolution and detail of the collected data, the whole process is designed to be conducted blindly and anonymously, in accordance with standing new European personal data protection laws (i.e. GDPR).

The second important pillar of the design of the application is driver feedback. By developing innovative measures within the smartphone application, and the respective web-platform, it is sought to raise the awareness of drivers with regards to their safety performance, and ultimately to motivate them to improve their driving behavior. Dedicated fields regarding driver errors, crash risk and skill improvement are designed. The developed experimental protocol includes feedback application on two levels (a) personalised feedback will be communicated to all drivers, with statistics and reports, allowing them to identify their critical deficits or unsafe behaviors for different modes, and (b) incentives within a social gamification scheme, with personalised target setting, benchmarking and comparison with peers.

To determine whether the provision of driver feedback has an impact on driving behavior, a series of analytic techniques will be applied on the data provided by the application. Already during the familiarization phase of the experiment (phase 1), which is spans its first two months, descriptive analyses will be conducted on the metrics of each driver to profile them in comparison to the sample average. The naturalistic experiment is considered suitable for this purpose, as this method is known to enable profiling concerning more aggressive driving behaviors that are in turn also related to road crashes (Tselentis et al., 2017). Statistical models will be then used to provide refined insights on the performance of drivers, and to determine statistically significant factors that impact it. Throughout the development process, the relevant literature has been reviewed, and key points have been summarized both from existing scientific sources (e.g. Regan et al., 2013) and from the review process carried within the project (Kontaxi et al., 2019).

At the time of writing of the present study (September 2019), 132 car drivers, 27 motorcycle drivers and 2 bicycle drivers have been recruited and drive their vehicles daily with the BeSmart application, with the recruitment process being open for more participants (professional driver groups are expected to engage soon as well).

3. Theoretical background

3.1. k-means Clustering

As a preliminary step, clustering is useful in order to divide the trip sample into several categories, which can provide insights as to whether driving behavior differs systematically on a macroscopic scale. K-means clustering is a very well-known and straightforward algorithm used to separate datasets into clusters, and belongs to the unsupervised machine learning algorithms. The algorithm searches for a specific number (k) of clusters in a dataset. The algorithm first initiates by randomly selecting centroids in the data. Each data-point is then assigned to the nearest centroid, forming k clusters. Centroids are re-computed for the formed clusters, and thus their location changes. Calculations are then performed to re-assign each data-point to their new centroid. Afterwards, iterative calculations are conducted until no reassignments are made and thus the centroids have stabilized. The popularity of k-means algorithms presented in the past (e.g. Hartigan & Wong, 1979) has led to several customized approaches in the literature (e.g. Kanungo et al., 2002; Likas et al., 2003). K-means has been used widely for clustering purposes in several transport/road safety studies as well (e.g. Yannis et al., 2007; Mantouka et al., 2019).

There is a number of methods to determine the optimal number of clusters for any dataset; in the current study, the elbow method is followed, which defines clusters by minimizing the total intra-cluster variation (expressed by the within cluster sums of squares (WCSS) (Kodinariya & Makwana, 2013).

3.2. Generalized Linear Mixed-Effects Models

This analysis aims to examine factors affecting event counts across drivers (i.e. the number of harsh acceleration and harsh brakings per trip). If one considers harsh events as instances similar to road crashes, but more frequent, utilizing Generalized Linear Models (GLMs) which are used when dealing with count data (Lord & Mannering, 2010) becomes meaningful. Furthermore, it can be reasonably assumed that each driver has their own traits such as personality and experience, which affects their driving style, and thus the counts of events they exhibit. In that sense, multiple trips cannot be considered as independent inputs, which violates the independence assumption common for linear models (Winter, 2013).

Therefore, in order to capture personal driver traits that are unobserved, random effects are introduced to GLMs in order to extend them as Generalized Linear Mixed-Effects Models (GLMMs). In other words, GLMMs are formulated by using the fixed effects GLMs as basis and then adding random effects to the linear equation. GLMMs were introduced by Breslow & Clayton (1993).

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

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

And the linear predictor is:

$$log(\lambda_i) = \beta_0 + \beta_n x_n + \varepsilon \tag{2}$$

Where β are the fixed-effect parameters (constant and coefficients) for n independent variables, and ε is the error term. The GLM is then extended into a GLMM by adding random effects. 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:

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

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

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

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

Coefficient result interpretation is more intuitive when using relative risk ratios (sometimes called incidence rate ratios). Relative risk ratios are obtained by transforming the predictor to obtain the frequency. For an increase of one unit in one specific variable, k, with all other parameters remaining equal, the predicted original frequency λ_i is multiplied by: $\lambda_{ki} = \exp(\beta_{ki}) * \ \lambda_i$

For computational reasons during the GLMM fitting, the trip data underwent z-score scaling, a common standardization process which does not affect the obtained coefficients. Mathematically, for every parameter x with a mean $\bar{x}$ and a standard deviation S a scaled value is obtained:

$$x_{scaled} = (x - \bar{x})/S \tag{6}$$

The best-fitting model which contains the more informative variable combination and explains the highest degree of variance per given dataset is selected as the one with the minimum Akaike Information Criterion (AIC). It is critical to note that the added value of any random effects is assessed by conducting a custom ANOVA (log-likelihood test) between the fixed effects GLM and any formulated GLMMs.

4. Analysis and Discussion

4.1. Descriptive statistics

Overall, 11211 trips from 132 car drivers were recorded using the BeSmart application, and these comprise the sample analyzed in this study. The trips amount to 11215302 seconds of total driving time (or about 130 days), and to 116934 total vehicle kilometers. Furthermore, a total of 8295 harsh acceleration events and 15821 harsh braking events were recorded in these trips, with varying frequencies per trip, as shown in Figure 1a and 1b.

All trips are currently from the familiarization phase (phase 1), which entails the application functioning in a passive recording state. This means that drivers have received no feedback on their driving performance yet; the following is a snapshot of the baseline driver behavior. R-studio has been used (R Core Team, 2019) for the analyses. Due to ease in certain aspects of the data extraction process, some variables were more readily obtainable for the analysis. Their descriptive statistics per trip are shown on Table 1.

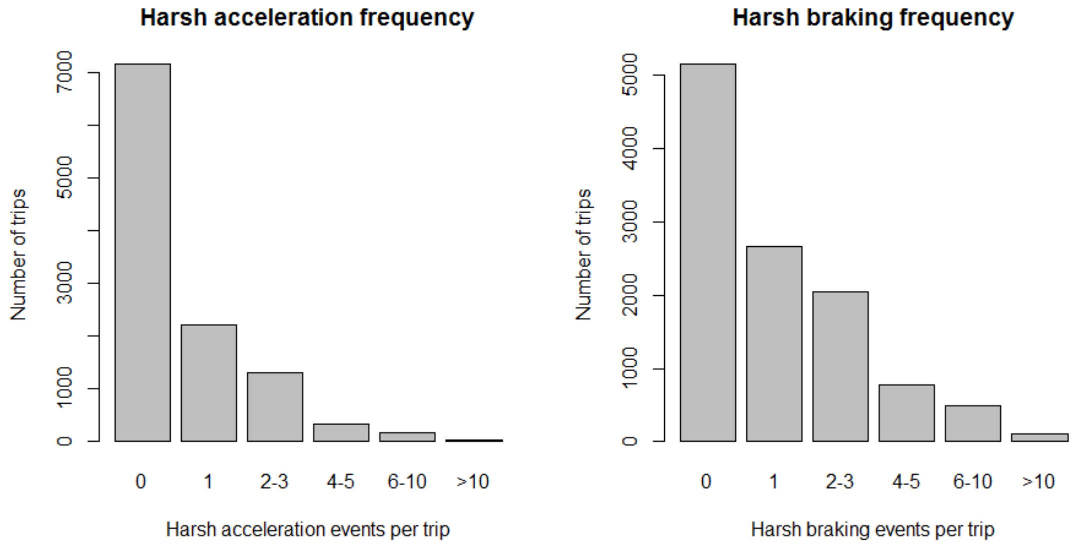

Figure 1: (a) Harsh acceleration frequencies per trip; (b) Harsh braking frequencies per trip

Table 1. Descriptive statistics of variables from 11211 trips of 132 car drivers.

Variable	Minimum	Mean	Maximum	St. Dev.
Total Trip Duration [s]	61.00	1000.38	17642	1040.50
Total Trip Distance [km]	0.50	10.43	387.20	19.89
Average Speed [km/h]	14.00	37.98	133	15.40
Maximum Speed [km/h]	19.00	69.43	206.00	26.35
Mobile Use Duration [s]	0.00	44.25	2719.00	148.77
Speeding Duration [s]	0.00	74.81	4747.00	204.76
Harsh accelerations [count]	0.00	0.74	19.00	1.46
Harsh brakings [count]	0.00	1.41	33.00	2.18

4.2. k-means Clustering

After establishing the theoretical background, several clustering configurations were tested, utilizing the algorithm of Hartigan and Wong (1979) for k-means clustering (default for R-studio). Ultimately, it was determined that more informative clusters could be obtained after eliminating outlier trips with exceedingly high amounts of events. Trips with more than 7 harsh braking or 7 harsh acceleration events (separately, not additively) were removed from the clustering dataset (amounting to 322 trips, or a 2.87% reduction).

Based on the results of the Elbow Method, which are shown in Figure 2a, 3 clusters were selected. The clusters appear on Figure 2b visually in a 2-D rendition. From the cluster analysis it can be determined that the highest degree of variability amongst trips is explained by fluctuations in maximum speed reached during the trip and the amount of harsh brakings done by drivers. Furthermore, the centroid values of clusters are shown on Table 2.

Table 2. Centroid centers and types for trip clusters

Nr	Trip Type	Total Duration	Total Distance	Av. Speed	Max. Speed	Mob. Use Dur.	Speeding Dur.	Harsh accel.	Harsh brak.	Speeding Dur./ Tot. Dur.	H.acc/ Total Distance	H.brk/ Total Distance
1	Aggressive	510.21	4.11	34.09	60.39	20.25	26.24	0.38	0.74	0.05	0.09	0.18
2	Speeding	4509.47	72.82	62.34	104.54	215.47	520.91	1.58	2.95	0.12	0.02	0.04
3	Average	1747.32	17.18	44.01	85.93	78.76	119.56	1.15	2.13	0.07	0.07	0.12

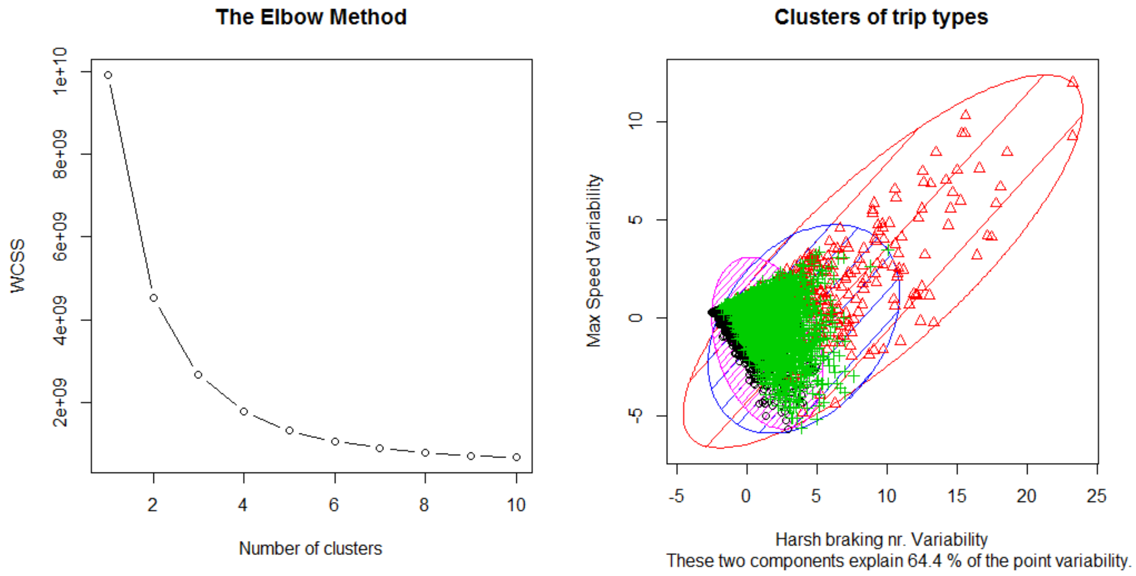

Figure 2: (a) Elbow Method for cluster number determination; (b) 2-D Cluster visualization

From the interpretation of the centroid values of clusters, three different trip types can be determined:

- Cluster 1 can be considered as the one comprising aggressive trips. These trips are short in duration and distance, but exhibit large amounts of harsh events per distance covered. They probably involve short local trips in familiar areas, in which the drivers are inattentive or distracted, thus cause more harsh events.
- Cluster 2 can be considered as the one comprising speeding trips. These trips have large duration and distance, and while they have low harsh event counts, they have very large durations of speeding. They are likely highway trips.
- Cluster 3 can be considered as the one comprising average trips. These trips have some harsh events but their overall traits are within normal values.

Mobile use duration (as a percentage of trip duration) appears to be overall evenly distributed among all trip types.

4.3. Generalized Linear Mixed-Effects Models

In order to model the expected number of events per trip for the participant drivers, models in a GLM framework were calibrated, as previously explained. Since the BeSmart application allows for a high resolution, big-data oriented collection scheme, it was attempted to include random effects in order to capture the unique driving behavior traits for each driver. This entails having a critical minimum sample of trips for each driver to achieve a meaningful outcome. Therefore, a screening was made among participant drivers, and drivers that had over 40 trips each were selected for the GLMM analysis.

From the 132 car drivers, 76 were ultimately selected. This was cumulative to the harsh event outlier removal described previously. While this might seem like a considerable loss of driver diversity, only 728 trips in total were removed from the initial dataset (resulting in a 6.94% reduction). GLMMs were then 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 Poisson function with the log odds link function was implemented. It should be noted that conceptually for harsh event counts, random slope models without random intercepts make little sense. After conducting log-likelihood test (ANOVA) comparisons, the most informative configuration of random effects was the inclusion of both random intercepts and random slopes in the GLMMs to capture unique driver traits (lowest LogLikelihood and highest χ^2). Results of mixed effect selection are shown on Table 3 for harsh accelerations and on Table 4 for harsh brakings:

Table 3. Log-likelihood comparison of mixed effect selection for harsh acceleration models

Model Family	Model Configuration	LogLikelihood	D.f.	χ^2	P(>χ^2)	Sig.
GLM	Fixed effects only [baseline]	-4844.60	6	–	–	–
GLMM	Fixed effects & Random Intercepts	-4262.90	7	1163.38	<2e-16	***
GLMM	Fixed effects, Random Intercepts & Random Slopes	-4224.00	9	77.78	<2e-16	***

Significance codes: '***': 0.000 \| '**': 0.001 \| '*': 0.01 \| '.': 0.05 \| ' ': ≥ 0.1

Table 4. Log-likelihood comparison of mixed effect selection for harsh braking models

Model Family	Model Configuration	LogLikelihood	D.f.	χ^2	P(>χ^2)	Sig.
GLM	Fixed effects only [baseline]	-6486.50	5	–	–	–
GLMM	Fixed effects & Random Intercepts	-5635.50	6	1702.02	<2e-16	***
GLMM	Fixed effects, Random Intercepts & Random Slopes	-5577.60	8	115.77	<2e-16	***

Significance codes: '***': 0.000 \| '**': 0.001 \| '*': 0.01 \| '.': 0.05 \| ' ': ≥ 0.1

The final models were selected as the ones with the lowest AIC values. Fixed effect results appear on Table 5 for harsh acceleration counts and for harsh braking counts:

Table 5. GLMMs for harsh event counts of 76 drivers (fixed effects).

Parameter	GLMM for harsh acceleration counts					GLMM for harsh braking counts				
	Estimate	s.e.	p-value	Sig.	Relative Risk Ratio	Estimate	s.e.	p-value	Sig.	Relative Risk Ratio
Intercept	-0.094	0.098	0.335		0.910	0.655	0.078	0.000	***	1.925
Maximum Speed	0.218	0.029	0.000	***	1.244	0.203	0.021	0.000	***	1.225
Speeding Duration	0.032	0.009	0.000	***	1.033	0.031	0.007	0.000	***	1.031
Mobile Use Duration	0.018	0.008	0.032	*	1.018	–	–	–	–	–
Log(Total Trip Duration)	0.291	0.042	0.000	***	1.338	0.280	0.030	0.000	***	1.324
Log(Total Trip Distance)	-0.066	0.022	0.002	**	0.936	-0.072	0.016	0.000	***	0.931

Significance codes: '***': 0.000 \| '**': 0.001 \| '*': 0.01 \| '.': 0.05 \| ' ': ≥ 0.1

The visual representations of values of random intercepts and random slopes for the log of total trip duration per driver are shown in Figure 3 for harsh acceleration counts and in Figure 4 for harsh braking counts, respectively. Personal differences per driver from the fixed effect intercept and slope are thus included in the linear predictor.

GLMM results indicate a number of correlations: The parameters of maximum speed, speeding duration and total trip duration have all been determined as statistically significant and positively correlated with both harsh acceleration and harsh event counts. Similarly, total trip distance is statistically significant and negatively correlated with both harsh event types. Mobile use duration was found significant only for harsh accelerations with a small positive correlation.

These correlations can be interpreted and quantified further using the relative risk ratio: An increase of the maximum speed of a trip by 1 km/h creates 1.24 times the harsh acceleration and 1.22 times the harsh braking events (or, increases their numbers by 24% and 22%, equivalently). Increases of 1s in speeding duration leads to about 3% increases in harsh event counts for both types. Similarly, increases of 1s in mobile use duration leads to about 2% increases in harsh accelerations. These effects are seemingly smaller, until one considers that the increases are multiplicative, and thus long durations of these violations create significant numbers of harsh events as well. All these results are expected, given the well-known adverse effects of speeding and distracted driving in road safety. The fact that mobile phone use was not found significant for harsh braking frequency is explained by overcompensation effects: drivers realize that they are distracted as they talk on their phones, and decelerate in order to compensate for the distraction. This leads to lower speeds and increased headways in instances when they might have harshly braked normally.

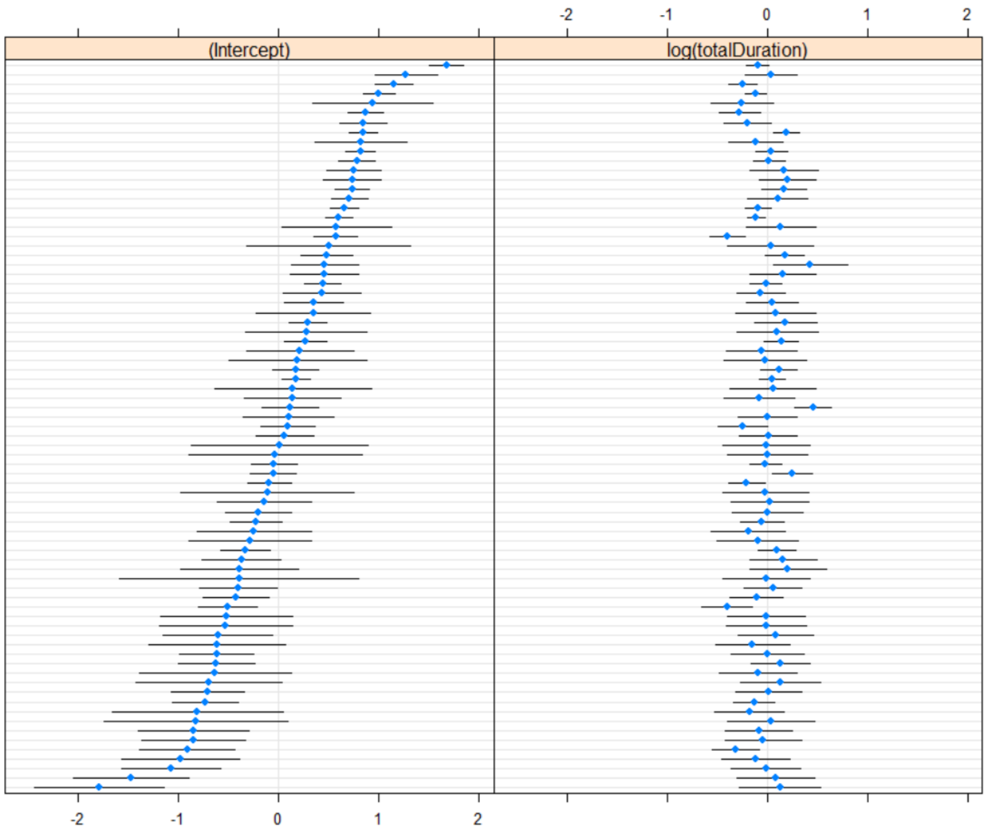

Figure 3: (a) Random Intercepts and (b) Random Slopes for log(total trip duration)
[Both in the GLMM for harsh acceleration counts]

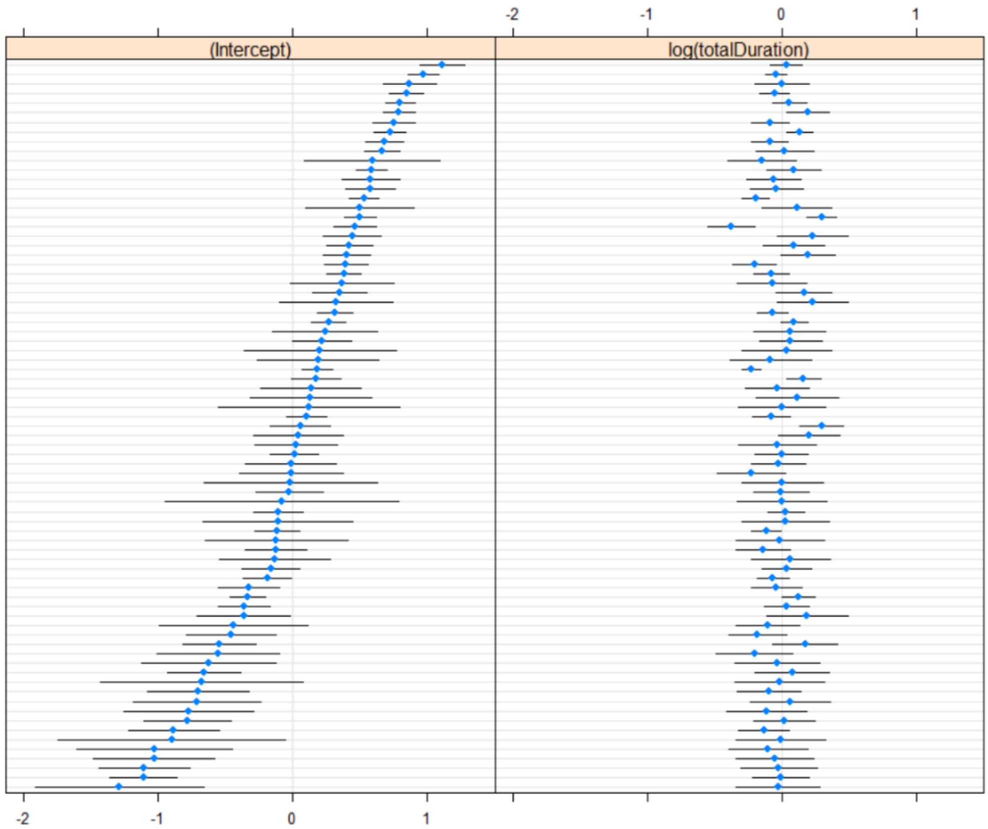

Figure 4: (a) Random Intercepts and (b) Random Slopes for log(total trip duration)
[Both in the GLMM for harsh braking counts]

The exposure parameters (trip distance and duration) were found to fit better in the model through a logarithmic relationship, which is a common practice to avoid negative projections. Increases in total trip duration also appear to lead to increases in harsh events. This can be interpreted as an exposure factor: after a point, drivers experience fatigue or impatience that leads to mistakes or encounter other drivers with behavior which is very different than their own; it is probable that harsh events occur then.

Total trip distance appears to be the only parameter that is negatively correlated with harsh event frequency. This finding hints at a possible relationship with driver mentality. Drivers that know that they have a large distance to cover are more prepared, physically and psychologically, and possibly more careful. They can understand that aggressive behavior spanning seconds or less will not affect the outcome of their trip positively, nor will it reduce travel time by any measure. Conversely, drivers that have a short distance to cover might have an urgency to complete their trip in order to do other tasks.

It is important to note that almost all parameters, and especially exposure parameters, had almost identical effects on harsh braking and harsh acceleration types. Although they are different phenomena, further corroborated by the findings for mobile use, they could be analyzed collectively as well.

5. Results of the present research

The current research aimed to showcase the potential of BeSmart for data collecting and road safety analyses. To that end, an analysis of car driver trips made during the familiarization phase of BeSmart was conducted. 11211 trips from 132 car drivers were examined and clustered in three types using k-means clustering after outlier removal: Aggressive, Speeding and Average trips. Generalized Linear Mixed-Effects Models were fitted to the trips of 76 car drivers who made frequent trips (>40 trips during the familiarization phase) in order to model the frequencies of harsh acceleration and harsh braking events. These events are registered in the application when acceleration over a certain threshold is recorded.

After log-likelihood test (ANOVA) comparisons, it was determined that additional random intercepts and random slopes in the model explained some of the variance in harsh event frequency caused by unobserved driver traits. Results indicate that maximum speed, speeding duration and total trip duration are positively correlated with both harsh acceleration and harsh event frequencies. Total trip distance was found to be statistically significant and negatively correlated with both harsh event types. Mobile use duration was found significant only for harsh accelerations with a small positive correlation.

6. Future Results from the BeSmart Project

As evident from the analysis of car drivers of the initial phase of BeSmart, multi-dimensional results are expected from the project experiment. First of all, the demands of a driver monitoring application through smartphone data collection are constantly determined and revisited, along with the related costs. In addition, high-resolution data will enable precise monitoring of driving behavior, and several variables will be examined. No other equipment or vehicle instrumentation is required for the functions of the application, which allows for an easy, compact and transferable data collection scheme.

One of the most important contributions that is expected from the BeSmart project is the precise and quantitative documentation of the impact of driver feedback through the personalized smartphone application. This feedback is expected to trigger drivers' learning and self-assessment process and enable them to gradually improve their performance and monitor their evolution. Success on the 200-driver sample will increase expectation for more scalable results in the future, with much more wide-spread road safety improvement effects, provided through dedicated smartphone applications.

A series of future analyses will be undertaken as the experiment progresses and drivers receive feedback on their performance. The initial planning involves time-series analyses, which will be employed to monitor temporal individual and group performance through stratification per time unit. Moreover, analyses per gender, age and primary transport mode will be explored in order to capture any particular trends found in the categories of these parameters, possibly improving feedback processes. An approach that is currently being examined involves comparing specific driver datasets across the experiment phases while controlling for their exposure, in order to capture the influence of every feedback feature that is provided.

The BeSmart project is driven by a user-centred, ergonomic and low-cost instrumentation approach, which allows flexibility in data collection and handling, and increases the exploitation potential for wide acceptance by the end users of the proposed tools and the respective stakeholders. Larger-scale impacts include significant benefits on road safety and other societal impacts, new approaches for driver training and support, as well as safer use of vehicles overall. Updates on the phases of the experiment and the overall progress and results of the project can be found at the BeSmart website: https://besmart-project.gr/.

7. 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 behavior and safety support system on the basis of smartphone applications.

References

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.

Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. Accident Analysis & Prevention, 123, 114-122.

Guo, F., Klauer, S. G., Fang, Y., Hankey, J. M., Antin, J. F., Perez, M. A., ... & Dingus, T. A. (2017). The effects of age on crash risk associated with driver distraction. International journal of epidemiology, 46(1), 258-265.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. Applied Statistics, 28, 100–108. doi: 10.2307/2346830.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), 881-892.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

Kontaxi A., Ziakopoulos A., Tselentis D., Yannis G. (2019). "A review of the impact of driver distraction on driving behavior and road safety." 9th International Congress on Transportation Research (ICTR) 2019, Athens, 24-25 October 2019.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.

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.

Ma, Y. L., Zhu, X., Hu, X., & Chiu, Y. C. (2018). The use of context-sensitive insurance telematics data in auto insurance rate making. Transportation Research Part A: Policy and Practice, 113, 243-258.

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. Safety Science.

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.

Nikias, V. A., Vlahogianni, E. I., Lee, T. C., & Golias, J. C. (2012). Determinants of powered two-wheelers virtual lane width in urban arterials. In Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on (pp. 1205-1210). IEEE

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Regan, M. A., Lee, J. D., & Victor, T. W. (2013). Driver Distraction and Inattention: Advances in Research and Countermeasures (Human Factors in Road and Rail Transport). Ashgate Publishing Group.

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. Accident Analysis & Prevention, 43(5), 1771-1781.

Theofilatos, A., Ziakopoulos, A., Papadimitriou, E., & Yannis, G. (2018). How many crashes are caused by driver interaction with passengers? A meta-analysis approach. Journal of safety research, 65, 11-20.

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.

WHO: Global status report on road safety by World Health Organization, 2018

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In Icml (Vol. 1, pp. 577-584).

Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. arXiv preprint arXiv:1308.5499.

Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modelling for the regional effect of enforcement on road accidents. Accident Analysis & Prevention, 39(4), 818-825.

Zaldivar, J., Calafate, C. T., Cano, J. C., & Manzoni, P. (2011). Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In Local Computer Networks (LCN), 2011 IEEE 36th Conference on (pp. 813-819). IEEE.