Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2022) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Transport Research Arena (TRA) Conference

Establishing the relationship between crashes and unsafe driver behaviors in motorway segments

Aikaterini Deliali[a*], Apostolos Ziakopoulos[a], Anastasios Dragomanovits[a], Ioannis Handanos[b], Christos Karadimas[b], George Kostoulas[c], Eleni Konstantina Frantzola[c], George Yannis[a]

[a]*National Technical University of Athens,5 Iroon Polytechniou, Athens, 15773, Greece*
[b]*Olympia Odos Operation SA, Vlichada 19100, Greece*
[c]*OSeven, 27B Chaimanta str., Chalandri 15234, Greece*

Abstract

Surrogate safety analysis is an alternative approach to crash-based analysis for the assessment of traffic safety. Surrogate safety analysis relies in various data types, and it is well admitted that the advancement of technology has enabled the recording of various types of traffic data (e.g., vehicle trajectories) that can be used for safety analysis. However, it is first needed to ensure that these data are correlated with crash data and so, can be used as a proxy of crash data. This study develops crash prediction models for motorway segments using harsh braking and harsh acceleration, and speeding data recorded via a smartphone app. The models indicate that harsh acceleration events are a good predictor of average crash frequency.

Keywords: crash prediction; driver behaviour; harsh braking; speeding; crowd-sourced data; surrogate safety.

1. Introduction

Crash occurrence analysis role in road safety management process is critical. However, the analysis of crash occurrence data is limited by crash availability or reliability, while at the same time crash analysis is a reactive approach in the sense that crashes need to occur before the consideration of safety countermeasures. Proactive or surrogate (i.e., non-crash based) safety approaches have been gaining popularity, a trend affected by technological advancements. Surrogate safety approaches use various metrics to assess safety which all describe unsafe traffic events

* *Aikaterini Deliali: +30 6971551576*
kdeliali@mail.ntua.gr

such as speeding, traffic conflicts, harsh braking, and others (Tarko, 2018). These metrics can be obtained through real world traffic data, e.g., GPS devices, smart phone applications (apps).

In the surrogate safety literature, a lot of attention has been placed on the comparison of metrics to identify traffic conflicts (e.g., time to collision or post encroachment time) and on the validation of surrogate safety metrics, correlate those metrics to crash occurrence. Technological developments in naturalistic driver recording and monitoring allowed for more seamless measurement of harsh driving events in addition to speeding events. Two types of harsh driving events can be denoted, namely harsh braking (HB) and harsh acceleration (HA), as events of abrupt and unexpected driving behavior.

Up to date, it remains unclear how other metrics, such as speeding or harsh braking, compared to each other in terms of effectiveness and reliability. There is no research or guidelines on determining how much data is needed for assessing safety based on surrogate safety metrics; at least three years of crash data are needed for conducting a reliable crash analysis. Overall, it is unclear which surrogate safety metrics are reliable for safety assessment (e.g., is speeding a more informative metric compared to harsh braking?). This paper aims to address these research gaps by establishing the relationship between crashes and three unsafe driver behaviors, namely speeding, harsh braking and harsh acceleration, in motorway segments. The outcome of this work will be a significant step towards proactive safety assessment while road authorities and traffic engineers will assess safety more efficiently as they will rely on fewer data, which can be collected easier.

The rest of the paper is organized as follows. Section two presents an overview of the literature on harsh braking, harsh acceleration, and speeding events. Section three presents the methodology and the data used for the analysis. Findings of the analysis are discussed in section four and lastly, the conclusions and next steps are in section five.

2. Background

Harsh braking events are generated by drivers as reaction to potentially hazardous situations (e.g., near-misses). Therefore, they are used as surrogate safety in naturalistic driving data and driving simulator studies (see for example Hanowski et al., 2005, Olson et al., 2009, Zohar et al., 2014, Petraki et al., 2020; Ziakopoulos et al., 2022). Relevant research has also documented harsh driving behavior as a component of driver profiling (Mantouka et al., 2019). Harsh braking events are among driver behavior parameters that are expected to be used in road safety modelling as intuitively they appear to be related with crash occurrence probability (Tselentis et al., 2017), however, there is no model up to date associating the occurrence of harsh braking events with crash occurrence.

Harsh acceleration events can be seen as a proxy of unsafe and/or aggressive driver behavior (Tselentis et al., 2017) and in turn, they can be seen as a proxy of traffic safety. Indicatively, drivers with higher anger, frustration and anxiety levels display higher acceleration values and apply increased physical pressure on the accelerator pedal (Stephens and Groeger, 2009). There are limited efforts up to date to link harsh acceleration events to crash occurrence. The research by Wåhlberg (2004) was inconclusive regarding the link between acceleration events and crash occurrence.

A wide body of the literature has been dedicated on exploring the relationship between speeds and crash occurrence and/or crash severity (see for example Elvik, 2013; Quddus, 2013; Aarts and Van Schagen, 2006; Cooper, 1997). The conclusions from these studies suggest that speed dispersion and differences in speeds are related to crash rates while before-after changes in speeds affect crash occurrence. Several studies have analyzed crash reports with the objective to relate crashes and speeding events and have revealed that speeding is not always correctly reported (Fitzpatrick et al., 2017; Cooper, 1997) and so, this is an indication that speeding should also be analyzed through other data sources instead of crash reports. The literature lacks associations between speeding events and crash occurrence.

3. Methodology

3.1. Model development

Regression models can be developed to predict the average crash frequency as a function of one or more independent variables. Crash frequency is commonly modelled using Poisson regression (Lord & Mannering, 2010). The probability of a segment i having y_i crashes per year (where y_i is a non-negative integer) is given by:

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \quad (1)$$

where $P(y_i)$ is the probability of a motorway segment i having y_i crashes per year and λ_i is the Poisson parameter for segment i, which is equal to the expected number of crashes per year, $E[y_i]$ for segment i. The Poisson parameter λ_i (expected number of crashes per period) needs to be specified as a function of independent variables. The most common functional form is:

$$\lambda_i = \exp(\beta X_i) \quad (2)$$

where X_i is a vector of explanatory variables and β is a vector of estimable parameters.

The Poisson regression assumes that variance and mean are equal, an assumption that usually does not hold for crash data. It has been found that crash data suffers from overdispersion (Lord & Mannering, 2010), meaning that a crash dataset (i.e., number of crashes over a period of years for several locations) has a mean that is lower than its variance. In practice this means that some locations (e.g., segments or intersections) aggregate more crashes than others. Therefore, for a dataset with a mean lower than its variance it is appropriate to use a negative binomial model (NB) instead of Poisson. In NB distribution, the variance varies from the mean by adding the term EXP(ε_i) (a gamma-distributed error term with mean 1 and variance) to the equation (2):

$$\lambda_i = \exp(\beta X_i + \varepsilon_\iota) \quad (3)$$

The form of the negative binomial distribution function is:

$$P(n_i) = \left[\frac{\frac{1}{\alpha}}{\frac{1}{\alpha}+\lambda_\iota}\right]^{(1/\alpha)} \times \frac{\Gamma\left[\left(\frac{1}{\alpha}\right)+n_i\right]}{\Gamma\left(\frac{1}{\alpha}\right)n_i\,!} \times \left[\frac{\lambda_i}{\left(\frac{1}{\alpha}\right)+\lambda_\iota}\right] \times n_i \quad (4)$$

where Γ. is a gamma function.

In this study four models were developed and compared. A base model was developed to predict the average crash frequency per segment based on traffic volume and length information, following the concept of the American Association of State Highway Transportation Officials (AASHTO) Highway Safety Manual (HSM) Predictive Method (AASHTO, 2010; 2014). The other three models aimed to examine whether the addition of more variables related to driver behavior can improve the performance of the crash prediction model. These models were:

- Base model + Harsh acceleration events
- Base model + Harsh braking events
- Base model + Speeding events

Models were developed using data from the Olympia Odos motorway in Greece; more details on the dataset are provided in subsection 3.2. For all five models, eighty percent of the dataset was used for model development while twenty percent of it was used for testing the model's performance.

3.2. Data

This study used data from the Olympia Odos motorway in Southern Greece. Olympia Odos connects two important cities of Greece, Athens and Patras. The motorway has a length of 201,5km, however, it operates in its full length since 2017. It has 29 interchanges and 25km of tunnels. For the analysis, four types of data were collected for the entire length of the Olympia Odos motorway: crash data, traffic data, road infrastructure data, and driver behavior data.

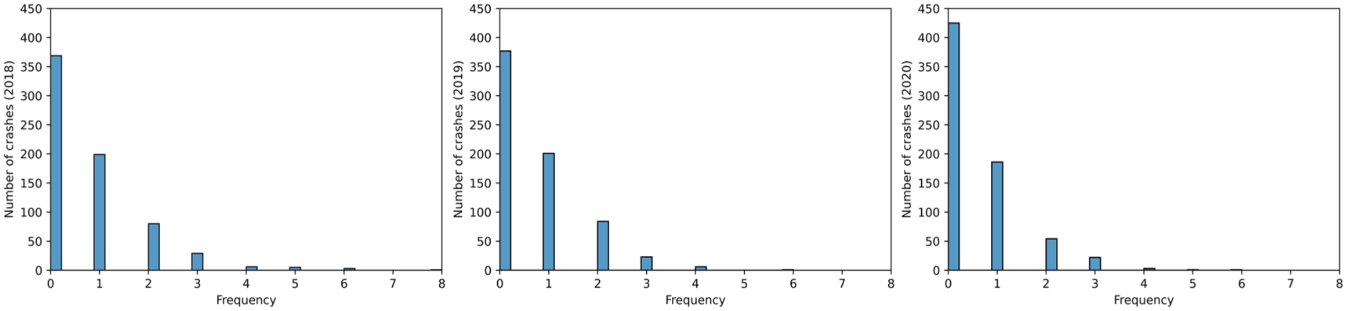
Crash data of all severity levels as well as Average Annual Daily Traffic data (AADT) were available for 2015-2020. As the entire motorway (i.e., from Athens to Patras) was finalized and started operating in 2017, AADT and crash data for the entire length were available for the years 2018-2020. It was selected to work with a smaller time period but for a longer road network. Summary statistics for crash traffic volume and segment length data are provided in Table 1, while frequency plots for the crashes are provided in Figure 1.

Data on horizontal curve radius, inside and outside shoulder widths, outside clearance, lane width, median width and barrier type, and interchange design characteristics were extracted using Google Maps imagery and Open GIS. Speed limit data was provided by Olympia Odos. All these data are available for each direction of traffic separately.

Table 1. Summary statistics for the crash and traffic data

Variable	Mean	St. Deviation	Min	Max
All crashes	0.66	0.65	0	4.33
Injury crashes	0.04	0.12	0	1
PDO crashes	0.62	0.63	0	4.33
Traffic volume (AADT)	10775	5524	6511	22078
Segment length	0.51	0.12	0.20	0.60

Naturalistic driver behavior data were recorded via a smartphone application (app) and processed in the platform,



both developed by OSeven. Drivers install the app developed by OSeven on their smartphone and subsequently engage in normal driving activities. The app engages automatically when driving is initiated, and records various data types such as vehicle location, speed, acceleration, deceleration, time/duration of engagement with the phone etc. These data are further processed to develop metrics to describe driver behavior. For this analysis, the following metrics for unsafe driver behaviors were used: counts of harsh accelerations and braking behaviors, average speed, average speed over the speed limit, count of trips with speeding, and average speeding duration.

OSeven has provided a representative dataset from its database that corresponds to the period from 01/6/2019 to 31/12/2020. The data provided by OSeven were in an anonymized format. A summary of these data for the years 2019 and 2020 is shown in Table 2. The data was recorded from a driver sample equal to 327 drivers 2019 and 330 drivers for 2020. For both years the total number of recorded trips is equal to 1.379.323.

Figure 1. Crash frequency for the years 2018, 2019 and 2020.

Table 2. Summary statistics for the representative dataset provided by OSeven.

	2019 (6 months)			2020 (12 months)		
	Harsh acceleration events	Harsh braking events	Speeding events	Harsh acceleration events	Harsh braking events	Speeding events
Mean	1.42	1.56	1.26	1.53	1.69	1.27
St. deviation	13.63	11.89	9.15	11.67	10.85	10.11
Variance	185.83	141.28	83.72	136.23	117.69	102.21

4. Results

The base model to predict average crash frequency for the motorway segments is presented in Table 3. Negative binomial distribution was used for the analysis as the mean and variance for the 2018-2020 crashes in Olympia Odos motorway were found equal to 1.98 and 3.92, respectively.

Both casualty and property damage-only (PDO) crashes included. The independent variables are the natural logarithm of AADT and length. Both of them were found positively related with crash frequency and statistically significant at the 95% confidence level. Model's performance was assessed using the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). For the based model it was found that AIC was equal to 1205.33 and BIC was equal to 1218.27.

Table 3. Base crash prediction model

Variable	**Coefficient**	**St. Error**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-6.9453	1.332	-5.214	0.000	-9.556	-4.334
ln(AADT)	0.7651	0.143	5.368	0.000	0.486	1.044
ln(length)	0.8366	0.265	3.156	0.002	0.317	1.356

The base model was extended to include an independent variable related to the presence of harsh acceleration (HA) events per motorway segment (Table 4). Harsh acceleration events data were used as percentage in the model: total number of recorded harsh acceleration events per segment over the total trips recorded per segment.

Table 4. Crash prediction model with harsh acceleration events

Variable	**Coefficient**	**St. Error**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-7.0100	1.337	-5.245	0.000	-9.630	-4.390
ln(AADT)	0.7740	0.143	5.415	0.000	0.494	1.054
ln(length)	0.9291	0.273	3.402	0.001	0.394	1.464
Average HA	20.1427	8.826	2.257	0.023	2.647	37.638

Similarly to AADT and length, harsh acceleration events are positively correlated with crash frequency and the variable is statistically significant at the 95% confidence level. AIC and BIC were equal to 1202.94 and 1220.20, respectively.

The base model was extended to include an independent variable related to the presence of harsh braking (HB) events. events per motorway segment (Table 5). Harsh braking events data were used in the same way as the harsh acceleration events, i.e., as percentage in the model: total number of recorded harsh braking events per segment over the total trips recorded per segment.

Table 5. Crash prediction model with harsh braking events

Variable	**Coefficient**	**St. Error**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-7.4137	1.335	-5.552	0.000	-10.031	-4.796
ln(AADT)	0.8134	0.143	5.690	0.000	0.533	1.094
ln(length)	0.8598	0.266	3.234	0.001	0.339	1.381
Average HB	18.4454	12.930	1.426	0.154	-6.899	43.788

In contrast to harsh acceleration events, the independent variable on harsh braking was not found to be statistically significant at the 95% confidence level. However, it is noteworthy that the sign of the coefficient is positive. Compared to the base model, the model that incorporated an independent variable on harsh braking events had AIC (1205.33) and BIC (1222.77) higher than the base model, which can be attributed to the fact that the additional variable was not significant.

The base model was extended to incorporate an independent variable on speeding events, as recorded by the smartphone app. As shown in Table 6, this new variable is not statistically significant at the 95% confidence level and does not improve the model's performance.

Table 6. Crash prediction model with speeding events

Variable	Coefficient	St. Error	z	P>|z|	[0.025	0.975]
Intercept	-7.2976	1.999	-3.651	0.000	-11.215	-3.380
ln(AADT)	0.8014	0.204	3.933	0.000	0.402	1.201
ln(length)	0.8085	0.262	3.082	0.002	0.294	1.323
Average Speeding	0.0289	0.646	0.045	0.964	-1.237	1.295

5. Conclusions

This study explored the relationship between crash occurrence and three unsafe traffic events, namely harsh acceleration, harsh braking and speeding in the case of rural motorways. Data on those events were available through a smartphone application (app), namely O-Seven (O7), that records vehicle kinematic data. Crash and traffic data for a three-year period were used to develop a "base" model to predict the average crash frequency for a motorway segment. The base model was extended to incorporate additional variables on harsh acceleration, harsh braking and speeding events and so, three new models were developed. The findings suggest that data on harsh acceleration events is positively correlated with the average crash frequency and has the potential to improve the performance of the base crash prediction model.

This study contributed to the surrogate safety literature by making a first attempt to associate crashes and unsafe driver events' data. While a direct relationship between crashes and unsafe traffic events was only found for the case of harsh acceleration events, it is recommended to explore whether the relationship between crashes and harsh braking events and crashes and speeding events remains the same in the case of other motorways. Future research should be extended to in general to other road settings, such as primary rural roads as well as in urban roads. Additionally, it is interesting to explore the amount of data needed to develop similar models.

Acknowledgements

This research is part of the project "i-safemodels – International Comparative Analyses of Road Traffic Statistics and Safety Modeling" of the National Strategic Reference Framework (NSRF): Greece – China Joint R&D Projects, funded by the "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) (2019- 2022).

References

AASHTO. 2010. Highway Safety Manual, 1st Edition, American Association of State Highway and Transportation Officials, Washington, DC, USA.

AASHTO. 2014. Highway Safety Manual, 1st Edition, American Association of State Highway and Transportation Officials, Washington, DC, USA

Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), pp.716-723.

Aarts, L. and Van Schagen, I., 2006. Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention, 38(2), pp.215-224.

Cooper, P.J., 1997. The relationship between speeding behaviour (as measured by violation convictions) and crash involvement. Journal of safety research, 28(2), pp.83-95.

Elvik, R., 2013. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis & Prevention, 50, pp.854-860.

Fitzpatrick, C.D., Rakasi, S. and Knodler Jr, M.A., 2017. An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression. Accident Analysis & Prevention, 98, pp.57-63.

Hanowski, R.J., Perez, M.A. and Dingus, T.A., 2005. Driver distraction in long-haul truck drivers. Transportation Research Part F: Traffic Psychology and Behaviour, 8(6), pp.441-458.

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.

Olson, R.L., Hanowski, R.J., Hickman, J.S. and Bocanegra, J., 2009. Driver distraction in commercial vehicle operations (No. FMCSA-RRT-09-042). United States. Department of Transportation. Federal Motor Carrier Safety Administration.

Petraki, V., Ziakopoulos, A. and Yannis, G., 2020. Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data. Accident Analysis & Prevention, 144, p.105657.

Quddus, M., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. Journal of Transportation Safety & Security, 5(1), pp.27-45.

Schwarz, G. 1978. Estimating the dimension of a model. The annals of statistics, pp.461-464.

Stephens, A.N. and Groeger, J.A., 2009. Situational specificity of trait influences on drivers' evaluations and driving behaviour. Transportation research part F: traffic psychology and behaviour, 12(1), pp.29-39.

Tarko, A. P. 2018. Surrogate measures of safety. In Safe mobility: challenges, methodology and solutions. Emerald Publishing Limited.

af Wåhlberg, A.E., 2004. The stability of driver acceleration behavior, and a replication of its relation to bus accidents. Accident Analysis & Prevention, 36(1), pp.83-92.

Ziakopoulos, A., Vlahogianni, E., Antoniou, C. and Yannis, G., 2022. Spatial predictions of harsh driving events using statistical and machine learning methods. Safety science, 150, p.105722.

Zohar, D., Huang, Y.H., Lee, J. and Robertson, M., 2014. A mediation model linking dispatcher leadership and work ownership with safety climate as predictors of truck driver safety performance. Accident Analysis & Prevention, 62, pp.17-25.