

SafeFITS - A GLOBAL MODEL AS A TOOL FOR ROAD SAFETY POLICY MAKING

George Yannis¹, Eleonora Papadimitriou², Katerina Folla³, Nenad Nikolic⁴, Eva Molnar⁵

Abstract: Road accidents constitute a major social problem in modern societies, with road traffic injuries being estimated to be the eighth leading cause of death globally making thus the need for action more and more pronounced. In this context, the SafeFITS, a global road safety model was developed for the United Nations Economic Commission for Europe (UNECE), which is based on global historical road safety data (72 indicators for 130 countries) and may serve as a road safety decision making tool for three types of policy analysis, i.e. intervention, benchmarking and forecasting analysis. For the development of the model, a hierarchical conceptual framework of five layers of the road safety system was designed (namely, economy and management, transport demand and exposure, road safety measures, road safety performance indicators, and road safety outcomes), and a dedicated database was developed with various road safety indicators for each layer. A two-step approach was opted for the purposes of the research, including the calculation of composite variables, their introduction in a regression model and the development of a model on the basis of short-term differences, accumulated to obtain medium- and long-term forecasts. The model developed has overall satisfactory performance and acceptable prediction errors, and preliminary validation provided encouraging results. Its usage might be proved highly useful for testing road safety policies, taking however into account the model limitations, mostly related to data availability and accuracy, and the recommendations for its optimal use.

Keywords: global road safety model, decision making tool, global database

¹ Professor, Yannis George, PhD Civil Engineer, National Technical University of Athens (NTUA) Iroon Polytechniou 9, Zografou 157 73, Athens, Greece, geyannis@central.ntua.gr

² Research Associate, Papadimitriou Eleonora, PhD Civil Engineer, National Technical University of Athens (NTUA) Iroon Polytechniou 9, Zografou 157 73, Athens, Greece, nopapadi@central.ntua.gr

³ Researcher, Folla Katerina, B.Sc. Civil Engineer, National Technical University of Athens (NTUA) Iroon Polytechniou 9, Zografou, Athens, Greece, katfolla@central.ntua.gr

⁴ Regional Advisor, Nikolic Nenad, Civil Engineer, UN Economic Commission for Europe, Sustainable Transport Division, Palais des Nations, Geneva, Switzerland, nenad.nikolic@unece.org

⁵ Former Director, Molnar Eva, UN Economic Commission for Europe, Sustainable Transport Division, Palais des Nations, Geneva, Switzerland, eva.molnar@unece.org

1. INTRODUCTION

Road accidents constitute a major social problem in modern societies, with road traffic injuries being estimated as the eighth leading cause of death globally and the first cause of death among young people aged between 15 and 29 years old. Particularly in low and middle income countries, road traffic injuries rates are twice those recorded in high income countries and still increasing, which is attributed at a large scale to the rapid motorization of many developing countries. Current trends suggest that, unless action is taken, the disparity between high- and low-income countries will further increase (WHO, 2015).

In order to guide countries on taking national-level actions to reduce road accident mortality, a Global Plan of Action was developed by the United Nations (UN, 2011), intended to serve as a guiding document for countries, and at the same time to promote coordinated action. Within this context, the UNECE launched the Safe Future Inland Transport Systems (SafeFITS) project, with aim to develop a road safety decision making tool for national and local governments both in developed and developing countries, to assist governments and decision makers to decide on the most appropriate road safety policies and measures in order to meet their targets.

This paper presents a global road safety model, developed within the SafeFITS project, which is based on global road safety data and may serve as a road safety decision making tool for three types of policy analysis, i.e. intervention, benchmarking and forecasting analysis. A conceptual framework of five layers of the road safety system was designed and a dedicated database was developed with various road safety indicators for each layer. A two-step modelling approach was implemented for the purposes of the research, including first the calculation of composite variables, and then their introduction in a generalized linear model correlating them with road safety outcomes.

2. MATERIAL AND METHODS

2.1. Methodological Framework

The methodological framework designed combines the five road safety pillars of Global Plan of Action (i.e. road safety management, road user, vehicle, road, post-crash care) (UN, 2011) with the concept of the SUNflower pyramid (Koornstra et al., 2002). As a result, the road safety management system is described as a hierarchical structure including the following five layers:

- The first layer, Economy & Management, reflects the structural, economic, cultural and regulatory characteristics of each country that are related to road safety performance.
- Transport demand and Exposure, at the second layer, reflects the characteristics of the transportation system and the exposure of the population due to urbanization, modal split, road network type, share of traffic per mode and per road type etc., which are all related to road risk.
- Road Safety Measures (policy output), at the third layer, are a result of structural and economical characteristics.

- To link these three layers to the actual road accident outcomes, an intermediate layer specifies the operational level of road safety in the country, containing road safety performance indicators (RSPIs).
- Final outcomes expressed in terms of road fatalities and injuries, which are necessary to understand the extent of the problem, comprise the fifth layer.

2.2. Modelling Approach

The main goal of the SafeFITS project was to develop a model that may serve as a tool for testing policy scenarios. In this context, it was necessary to take into account as many indicators as possible. In addition, efficient forecasting of future developments needs to take into account previous developments and therefore, make explicit consideration of the time dimension (Commandeur et al., 2013). In order to meet these requirements, a two-step approach was opted, which includes the calculation of composite variables (Al Haji, 2005; OECD, 2008; Bax, 2012) and then their introduction in a generalized linear model by taking into account past developments (namely on the basis of short-term differences, accumulated to obtain medium- and long-term forecasts).

Each layer of the modelling framework comprises numerous different indicators, from the five pillars of the UN Global Plan for Action (UN, 2011). In order to reduce the number of dimensions of the analysis, while exploiting as much information as possible, the analysis of composite variables (i.e. combinations of indicators), instead of individual indicators, was selected.

Overall, for a set of countries, (i) fatalities and injuries specific indicators were considered, (j) specific safety performance indicators, (k) road safety measures indicators, (l) transport demand and exposure indicators, and (m) economy and management indicators. Each layer is described by a composite variable (denoted as [Composite Variable]), estimated as a linear combination of several indicators through factor analysis as follows:

$$[Fatalities and Injuries] = \alpha_1 * (Fatalities and Injuries Indicator_1) + \alpha_2 * (Fatalities and Injuries Indicator_2) + \dots + \alpha_i * (Fatalities and Injuries Indicator_i) + e \quad (1a)$$

$$[RSPI] = \beta_1 * (RSPI Indicator_1) + \beta_2 * (RSPI Indicator_2) + \dots + \beta_j * (RSPI Indicator_j) + v \quad (1b)$$

$$[Road Safety Measures] = \gamma_1 * (Road Safety Measures Indicator_1) + \gamma_2 * (Road Safety Measures Indicator_2) + \dots + \gamma_k * (Road Safety Measures Indicator_k) + w \quad (1c)$$

$$[Transport demand & exposure] = \delta_1 * (Transport demand & exposure Indicator_1) + \dots + \delta_l * (Transport demand & exposure_l) + y \quad (1d)$$

$$[Economy & Management] = \varepsilon_1 * (Economy & Management Indicator_1) + \varepsilon_2 * (Economy & Management Indicator_2) + \dots + \varepsilon_m * (Economy & Management Indicator_m) + z \quad (1e),$$

with $\alpha, \beta, \gamma, \delta, \varepsilon$ parameters to be estimated and e, v, w, y, z error terms expressing the uncertainty in the estimation of the composite variables.

In the next step, the development of a model linking road safety outcomes with composite variables was pursued, in order to estimate the effect of individual indicators on road safety outcomes, through the composite variables. A logarithmic model is outlined as follows:

$$\text{Log}[\text{Fatalities \& Injuries}]_i = A_i + K_i * [\text{Economy \& Management}]_i + L_i * [\text{Transport demand \& Exposure}]_i + M_i * [\text{Road Safety Measures}]_i + N_i * [\text{RSPI}]_i + \varepsilon_i \quad (2)$$

with (i) countries, A , K , L , M , N parameters to be estimated, and ε error term expressing the uncertainty in the estimation of the relationship.

However, as mentioned above, the relationships between road safety outcomes and indicators depend on the underlying trends in the evolution of outcomes. The time dimension is taken into account by implementing a medium-term forecasting approach, on the basis of the developments over the last few years, for which data are available. By applying the same approach to the future forecasted outcomes, long-term forecasts are also obtained under certain conditions. The key variable that was taken into account in the forecasts to account for past and future developments is GDP, which is considered an appropriate indicator for forecasting road safety developments in the case of the absence of mobility and exposure data (Kopits and Cropper, 2005; Antoniou et al., 2016). Terms were introduced in the models, relating the road safety outcomes of year to those of previous years and to GDP (Yannis et al., 2014).

The final specification of the generalized linear model of Equation (2) including of short-term differences (τ years) in fatality rates is as follows:

$$\text{Log}(\text{Fatalities per Population})_{it} = A_i + \text{Log}(\text{Fatalities per Population})_{(t-\tau)} + B_i * \text{GDP}_{it} + K_i * [\text{Economy \& Management}]_{it} + L_i * [\text{Transport demand \& Exposure}]_{it} + M_i * [\text{Road Safety Measures}]_{it} + N_i * [\text{RSPI}]_{it} + \varepsilon_i \quad (3)$$

2.3. SafeFITS Database

The database was developed in order to cover the structure of the road safety management system as adopted in the context of the SafeFITS project. The relevant data were explored in international databases, such as World Health Organization (WHO), UN database, World Bank, International Road Federation (IRF) etc., aiming to collect reliable and most recent data for the greatest possible number of UN Member States. Consequently, data were collected for 130 countries (with population higher than 2,8 million inhabitants) and 72 indicators covering all layers. Data refer to 2013, for which there are the latest available fatality data.

An issue that was handled during the data preparation was the imputation of the missing values. First, for those variables and countries with available time-series, the latest available data were used for 2013. For the remaining countries, the substitution of the non-available data with the known mean value was selected. On that purpose, the countries were separated into three groups based on their motorization level, road safety performance and economic performance. Thus, the missing values of each indicator of the countries were filled with the known mean value of the indicator in the available group of countries in their regions.

3. RESULTS

3.1. Calculation of composite variables

The factor analyses were implemented on each one of the layers of the road safety, constrained to yield one factor per layer, an approach which lies within the family of “confirmatory” rather than exploratory factor analysis. The fatality rate per population was used as main dependent variable for two reasons, since there were no sufficient data for additional indicators of road safety outcomes to estimate a composite dependent variable.

Table 1 shows the factor loadings and score coefficients estimated by the confirmatory factor analysis of each one the indicators. Indicators with ‘loadings’ higher than 0.3 (which was the threshold set) were included in the calculation of the composite variables per layer:

- Economy and Management: indicators related to the demographic distribution are those with the highest loadings amongst, complemented with some elements of the road safety management system. This factor represents 34.7% of the overall variance in the data.
- Transport Demand and Exposure: indicators related to the vehicle fleet distribution are those with the highest loadings amongst, complemented with some elements of the road network and passenger/ freight kilometres in road transport. This factor represents 30.8% of the overall variance in the data.
- Measures: indicators related to the vehicle standards are the variables with the highest loadings amongst, followed by the blood alcohol concentration (BAC) limits, the speed limits and the measure on transport of dangerous goods (ADR). Several other indicators are included with lower loadings. This factor represents 34.2% of the overall variance in the data.
- Safety Performance indicators: all indicators tested had a high loading, bringing together all the elements of enforcement, as well as variables related to the use of safety equipment and post-impact care This factor represents 58.2% of the overall variance in the data.

Table 1. Indicator loadings and coefficients on the estimated factor (composite variable) per layer

		Factor (composite variable)	
		Loadings	Score coefficients
Economy and Management	EM2_lt15yo	-0.778	-0.250
	EM3_gt65yo	0.714	0.229
	EM4_UrbanPop	0.709	0.228
	EM7_NationalStrategy	0.697	0.224
	EM8_NationalStrategyFunded	0.626	0.201
	EM9_FatalityTargets	0.692	0.222
Transport Demand and Exposure	TE1_RoadNetworkDensity	0.497	0.161
	TE2_Motorways	0.460	0.149
	TE3_PavedRoads	0.734	0.238
	TE4_VehiclesPerPop	0.839	0.272
	TE5_PassCars	0.825	0.267

		Factor (composite variable)	
		Loadings	Score coefficients
	TE7_PTW	-0.681	-0.221
	TE10_PassengerFreight	-0.360	-0.117
Road Safety Measures	ME2_ADR	0.681	0.069
	ME4_SpeedLimits_urban	0.443	0.045
	ME6_SpeedLimits_motorways	0.634	0.064
	ME7_VehStand_seatbelts	0.877	0.088
	ME8_VehStand_SeatbeltAnchorage	0.906	0.091
	ME9_VehStand_FrontImpact	0.908	0.092
	ME10_VehStand_SideImpact	0.904	0.091
	ME11_VehStand_ESC	0.891	0.090
	ME12_VehStand_PedProtection	0.862	0.087
	ME13_VehStand_ChildSeats	0.896	0.090
	ME15_BAClimits	0.670	0.068
	ME16_BAClimits_young	0.670	0.068
	ME17_BAClimits_commercial	0.645	0.065
	ME19_SeatBeltLaw_all	0.570	0.057
	ME20_ChildRestraintLaw	0.628	0.063
	ME22_HelmetFastened	0.334	0.034
	ME23_HelmetStand	0.379	0.038
ME24_MobileLaw	0.375	0.038	
ME25_MobileLaw_handheld	0.350	0.035	
ME27_PenaltyPointSyst	0.378	0.038	
ME29_EmergTrain_nurses	0.399	0.040	
Road Safety Performance Indicators	PI1_SeatBeltLaw_enf	0.756	0.144
	PI2_DrinkDrivingLaw_enf	0.812	0.155
	PI3_SpeedLaw_enf	0.795	0.152
	PI4_HelmetLaw_enf	0.837	0.160
	PI5_SeatBelt_rates_front	0.811	0.155
	PI6_SeatBelt_rates_rear	0.766	0.146
	PI7_Helmet_rates_driver	0.784	0.150
	PI8_SI_ambulance	0.667	0.127
	PI9_HospitalBeds	0.607	0.116

3.2. Development of the Generalized Linear Model

Several alternative model specifications were tested for the selection of the final model. The best performing model for the purposes of the present research is presented in Table 2. The dependent variable is the logarithm of the fatality rate per population for 2013, and the main explanatory variables are the respective logarithm of fatality rate in 2010 (so the development of fatality rate over 2010-2013 is modelled), and the respective logarithm of GDP per capita for 2013, together with the four composite variables.

Table 2. Parameter estimates and fit of the final generalized linear model

Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	p-value
(Intercept)	1.694	0.2737	1.157	2.230	38.291	1	<0.001
Comp_ME	-0.135	0.0646	-0.261	-0.008	4.358	1	0.037
Comp_TE	-0.007	0.0028	-0.013	-0.002	7.230	1	0.007
Comp_PI	-0.007	0.0030	-0.013	-0.001	5.652	1	0.017
Comp_EM	0.007	0.0051	-0.003	0.017	2.009	1	0.156

LNfestim_2010	0.769	0.0462	0.678	0.859	276.322	1	<0.001
LNGNI_2013	-0.091	0.0314	-0.153	-0.030	8.402	1	0.004
(Scale)	0.038						
Likelihood Ratio	1379.00						
df	6						
p-value	<0.001						

The modelling results can be analysed as follows: An increase in the GNI results in decrease of the change in the fatality rate, which is in accordance to previous research findings. Additionally, a higher fatality rate in 2010 is associated with a higher fatality rate in 2013. This is also intuitive, as countries with higher fatality rates in the past are expected (all other things kept equal) to exhibit similar fatality rates in the future. All the parameter estimates of the composite variables on Measures or SPIs have a negative sign, suggesting that an increase in one or more of the indicators forming the composite variable results in a decrease in the fatality rate.

All the parameter estimates are statistically significant at 95% confidence level (p-values <0.050), and the Likelihood Ratio Test leads to accept the model, as its value is significant for an equal chi-square test with 6 degrees of freedom.

The quality of the model was also examined by comparing the observed and the predicted fatality rates. The mean absolute prediction error was estimated at 2.7 fatalities per population, whereas the mean percentage prediction error was estimated at 15% of the observed value.

A cross-validation was also carried out with two subsets of the sample. A randomly selected 80% of the sample was used to develop the model, which was implemented to predict the fatality rate of the 20% of the sample not used to fit the model. Similarly, a 70% of the sample was randomly selected to develop the model, which was fitted to the 20% of the remaining sample. In the first case, the mean percentage prediction error was 12% and in the second case the mean percentage prediction error was 19%, which makes the model of satisfactory performance.

3.3. Model application for Serbia

An example of the model application in Serbia is presented in this section. In the first step of the model application, the fatality rates per 100.000 population were forecasted, based solely on the projected changes of GNI per capita and demographic indicators (base case scenario). In this case, it was estimated that in 2030 the fatality rate in Serbia will be 11.03. In the next step, three sets of interventions were tested for 2022 and the fatality rates of each set are shown in Figure 1 alongside with the respective confidence intervals.

The first set of interventions includes the increase of seat-belt law enforcement from 6 to 9 (on a scale from 0 to 10), the increase of helmet law enforcement from 8 to 9, the increase of the seat-belt use rates in front seats from 65.8% to 80%, in the rear seats from 3.1% to 25% and the increase of the helmet use rates from 60 to 78%. The fatality rate for 2025 was estimated to be 10.02, which is 6.3% lower than the respectively estimated fatality rate in the base case scenario.

The second set includes, additionally to the first set, the implementation of a national road safety strategy, the increase of the percentage of motorways from 1.36% to 1.9%, increase of the percentage of paved roads from 66.19% to 85% and the introduction of

the ADR law. In this case, the fatality rate for 2025 was estimated at 9.63 which is 9.9% lower than the respectively estimated fatality rate in the base case scenario.

The third set of interventions includes, besides the aforementioned ones, the introduction in the national legislation four out of seven vehicle standards suggested by the UN (i.e. seat-belts, seat-belt anchorages, frontal impact and pedestrian protection). In this case, for 2025 9.26 fatalities per 100.000 population are estimated, 13.4% fewer than the forecasted for the same year in the base case scenario.

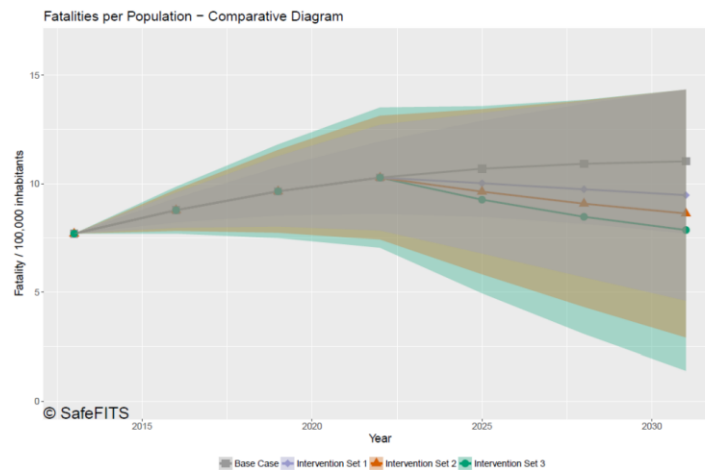


Figure 1. Forecasted fatality rates per 100.000 population in Serbia according to SafeFITS model

4. DISCUSSION

The developed model took into account several challenges and particularities of road safety analyses. The task of road safety forecasting on the basis of policy scenarios, i.e. combining an explanatory approach on road safety with the time dimension at global level, was a challenge on its own, as there was no similar example in the literature. The development of a dedicated methodology was required, different statistical techniques were combined and adjusted and several alternative hypotheses were tested, in order to meet the objectives of the analysis while dealing with data and methodological limitations.

While the model was developed on the basis of the most recent and good quality data available internationally, and by means of rigorous statistical methods, however, data and analysis methods have some limitations which should be kept in mind.

- The fatality data used for the model development are in some cases estimated numbers, and in all subject to under-reporting.
- There is lack of data, especially for transport demand, exposure and performance indicators and the missing values were replaced by the regional known mean value.
- In most cases, a binary variable (yes/no) was available, which may not always reflect the true value of the variable. For example, a measure may be

partially implemented, a national strategy may exist but there is no information whether it is implemented and monitored.

Consequently, the optimal use of the model depends on a number of recommendations and rules, in order to minimize errors and inaccuracies in the model outcomes, as follows:

- When used for forecasting purposes, the model can only be based on the extrapolation of short-term developments in the future; this approach has some obvious limitations. Confidence intervals for the predictions can be calculated to reflect the uncertainty in this extrapolation, on the basis of the mean prediction error of the model. The prediction error is considered to increase as the prediction horizon extends.
- The model includes many indicators which are correlated. However, composite variables may also be correlated with one another (e.g. measures with performance indicators), since correlation may exist between indicators included in separate composite variables. Therefore, the effects of interventions may not reflect the unique contribution of each separate intervention. When testing policy scenarios, it is strongly recommended to test combinations of “similar” interventions. The cumulative effect of “similar” variables indicators either within the same composite variable or from separate composite variables is more likely to accurately reflect true (and not conditional) effects.
- The model may not fully capture the effects on countries with very particular characteristics such as very low GDP, very high share of motorcycle or cyclist fatalities etc. Although every effort was made to customize the model for different geographical or geopolitical groups, as well as for such particularities, the available data in the international databases and the available information in the literature were not sufficient so far to allow for such customization.
- Developing countries are expected to be more sensitive in the testing of interventions than developed ones. There are several industrialized countries that already have very high values on all indicators, and their GDP is expected to keep increasing. For these countries, a further slightly decreasing trend is forecasted by the model, but in order to forecast substantial further reductions, other types of interventions will be required, for which no data is currently available. Therefore, the current forecasts for these countries may be quite conservative.

5. CONCLUSION

In the SafeFITS Project, a statistical model was developed on the basis of actual global road safety data, which can be used in three types of analyses, all very pertinent for road safety policy purposes, i.e. intervention, benchmarking and forecasting analysis. The proposed approach is based on the calculation of composite variables and their introduction in a generalized linear model (two-step approach), and forecast on the basis of short-term differences, accumulated to obtain medium- and long-term forecasts. Both these scientific choices have their limitations, but they were the optimal solutions for

dealing with the complexity of the model to be developed on the basis of the available data.

The final developed model is robust, with a satisfactory performance and acceptable prediction errors. The cross-validation undertaken is considered successful; however, care should be taken that the limitations of the model are taken into account, and several recommendations are made for optimal use of the model (e.g. combinations of policy scenarios).

The current model also has limitations related to data availability and accuracy. The lack of a global road safety database with detailed and comparable data certainly compromises the efforts to develop a global road safety model. Previous studies have indicated that there may be more data on exposure and SPIs at national level, than those reported in international statistics, and their collection, harmonization and use will be a major challenge with considerable added value for improving the model to better support road safety decision making.

Most importantly, a new wave of historical data may enable the time dimension within the model to be better taken into account, by estimating future developments on the basis of longer historical trends for fatalities and key economy, exposure and RSPI.. Additionally, further changes in programs and measures implemented in the various countries will allow to more accurately estimate their effects on outcomes, improving the transferability of estimates in other countries as well. It is therefore suggested to closely monitor global developments in data availability and accuracy, so that the data is updated regularly and continuously, allowing to improve the model with more, and more accurate data.

6. ACKNOWLEDGEMENTS

This paper is based on work carried out in the framework of the “Safe Future Inland Transport Systems (SafeFITS)” project of the United Nations Economic Commission for Europe (UNECE).

7. BIBLIOGRAPHY

- [1] Al Haji, G. (2005). Towards a Road Safety Development Index (RSDI) - Development of an International Index to Measure Road Safety Performance. Linköpings Studies in Science and Technology, Licentiate Thesis No. 1174, Norrköping, Sweden.
- [2] Antoniou C., Yannis G., Papadimitriou E., Lassarre S. (2016). Relating traffic fatalities to GDP in Europe on the long term, *Accident Analysis & Prevention*, Vol. 92, pp. 89-96.
- [3] Bax C. (Ed.), 2012. Developing a successful Composite Index; End report. Deliverable 4.9 of the Dakota project, European Commission, Brussels. (http://www.dacota-project.eu/Deliverables/DaCoTA_D4.9_developing%20a%20RSI%20deliverable.pdf).
- [4] Commandeur, J.J.F., F.D. Bijleveld, R. Bergel, C. Antoniou, G. Yannis, E. Papadimitriou (2013). On statistical inference in time series analysis of the evolution of road safety. *Accident Analysis and Prevention* 60, pp. 424-434.
- [5] Koornstra M., Lynam D., Nilsson G., Noordzij P., Pettersson H. E., Wegman F., Wouters P. (2002). *SUNflower - A comparative study of the development of road safety*

- in Sweden, the United Kingdom, and the Netherlands. Final report. SWOV Institute for Road Safety Research.
- [6] Kopits E., Cropper M. (2005). Traffic fatalities and economic growth. *Accident Analysis and Prevention* Vol. 37, pp. 169–178.
- [7] OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*, www.oecd.org/publishing/corrigenda, Organization for Economic Cooperation and Development, Paris.
- [8] UN (2011). *Global Plan for the Decade of Action for Road Safety 2011-2020*, United Nations (http://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf).
- [9] WHO (2015). *Global Status Report on Road Safety*. World Health Organization, Geneva.
- [10] Yannis G., Papadimitriou E., Folla K. (2014). Effect of GDP changes on road traffic fatalities. *Safety Science* Vol. 63, pp. 42-49.