Analyzing Road Safety Indicator Data across Europe: Describing, Explaining and Comparing

Brenda Wilmots, Elke Hermans, Tom Brijs and Geert Wets
Hasselt University
Transportation Research Institute - IMOB
Wetenschapspark 5 bus 6
3590 Diepenbeek
Belgium
E-mail: brenda.wilmots@uhasselt.be

ABSTRACT

Road safety is a relevant theme receiving increasing attention worldwide. Due to the various factors that influence the road safety level in a country, it is a complex topic to study. The use of indicators describing a particular aspect in an understandable way is valuable in this respect. Consequently, a diverse set of road safety related indicators (including road safety outcome indicators, transport indicators, demographic indicators, etc) is formulated in this study and data are collected for a large set of European countries. Next, this indicator set is analysed to develop the profile of a country, to compare the road safety performance between countries and to assess the explanatory power of the different indicators in terms of the number of road fatalities per million inhabitants. For each country, the aspects on which it performs (relatively) well and the aspects deserving further attention can be revealed. This is illustrated for Belgium in this paper. Moreover, countries with a similar (e.g. road safety outcome or transport) performance can be grouped. Finally, insight into the most determinant indicators for the number of road fatalities per million inhabitants offers useful information.

1. Introduction

Due to the human as well as financial suffering caused by road crashes, road safety is a relevant theme to study. The World Health Organization estimates that worldwide each year 1.3 million people are killed and between 20 and 50 million are injured in road crashes (WHO, 2009). In 25 countries of the European Union a total of almost 40,000 fatalities was registered in 2006, which is a decrease of 30% compared to the number in 1997. However, if the trend continues at the same rate, the European Commission's goal of halving the number of fatalities by 2010 will not be achieved (SafetyNet, 2008). Researchers like Sivak and Tsimhoni (2008) state that the easy problems in traffic safety have already been addressed (by taking measures based on common sense) and as a result more complex problems need to be handled to achieve further reductions in road crashes and casualties. To improve the road safety level in a country, it is therefore interesting to study the various factors that influence it, leading to valuable policy recommendations.

In this paper a set of road safety related indicators is developed and available road safety data are studied. Nowadays, the road safety level in a country is often described in terms of final outcome information, for example 'the number of killed and injured persons', using registered accident data (Hermans, 2009). Although these output indicators are relevant, they do not give a total view on the road safety situation in a country and the factors influencing it. Therefore, also road safety performance indicators and indicators related to the background and context of a country are taken into consideration in this research. The European Transport Safety Council (2001) defines a safety performance indicator as "any measurement that is causally related to crashes or injuries, used in addition to a count of crashes or injuries in order to indicate safety performance or understand the process that leads to crashes". Concerning the background and context of a country, demographic, economic, geographic, transport and mobility policy indicators are taken into account. For the formulation of relevant indicators literature has been consulted (for example Al Haji, 2005; European Transport Safety Council, 2001; Morsink et al., 2005; SafetyNet, 2005; Sartre 3 consortium, 2004).

By creating this diverse set of road safety related indicators and collecting indicator data for a large number of European countries, various analyses can be performed to describe, explain and compare European countries in terms of road safety. First of all, it is studied which of the indicators have the largest explanatory power in terms of the most often used road safety (output) indicator 'the number of road fatalities per million inhabitants'. That way, insight is gained into the influencing indicators. Secondly, clusters of homogeneous countries are identified in the data set. For different indicator sets, countries performing in a similar way are identified. Results are discussed for Belgium and the best performing countries with respect to road safety outcome (the so-called SUN-countries; Sweden, United Kingdom and the Netherlands). The two described analyses (regression and cluster analyses) are performed using the SPSS 17.0 software. Thirdly, the footprint methodology developed by Morsink et al. (2005) is illustrated. Based on this schematic overview of the relative performance, the good and bad aspects of Belgium are pointed out. Before presenting the results, the selected indicators and indicator data are further described. Next, the results of the analyses are discussed and in the last section the most important conclusions are summarized.

2. Indicators and data

Figure 1 gives an overview of the safety (final) output indicators, the safety performance indicators (SPI's) and the background and context indicators used in this paper. Each category is captured by means of a number of indicators.

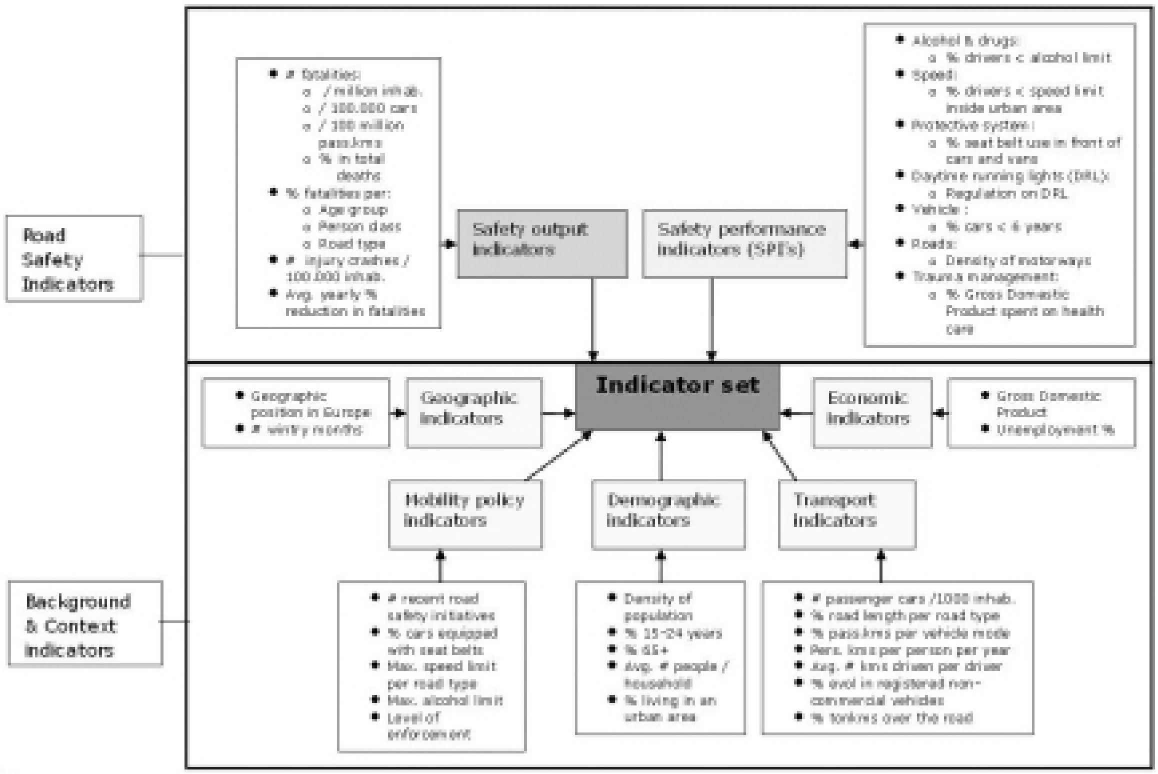


Figure 1 : Set of indicators used in this study

By means of published reports and online available databases (European Transport Safety Council, EuroStat, International Road Traffic and Accident Database, Organisation for Economic Co-operation and Development, etc), the indicators described above could be filled out for 21 European countries: Austria (AT), Belgium (BE), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IR), Italy (IT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SL), Spain (ES), Sweden (SE), Switzerland (CH) and United Kingdom (UK). Focus was on the most recent year for which data were available, being 2003. Nevertheless, some indicator values are missing for certain countries. The imputation of missing values is outside the scope of the current study. Therefore, only available data are used in the analyses. Furthermore, because the data set consists of different types of data, regression and cluster analyses are performed on the standardized indicator data set.

Oral Presentations

3. Method and Results

Next, the aim and methodology of regression, cluster and footprint analysis are briefly discussed. Emphasis lies on the results. We begin by identifying the key indicators that explain the number of road fatalities per million inhabitants. This dependent variable is chosen because it is the most common road safety output indicator and has reliable data available (Hermans, 2009).

3.1. Explaining the number of road fatalities per million inhabitants

3.1.1. Theoretical background and methodology

Malhotra & Birks (2003) describe regression analysis as “a statistical procedure for analysing the associative relations between a dependent variable and one or more independent variables”. These independent variables are included in the regression model to predict or explain the value of the dependent variable (Hair, 1998). In this paper, the variation in the number of road fatalities per million inhabitants (safety output indicator) that can be explained by the use of other types of indicators is assessed. The indicators or independent variables are added into a linear multiple regression model in various ways. A first approach first includes all variables in the model (using the Stepwise method) and only the ones having a high explanatory power remain. An alternative method incorporates the variables in different blocks according to the type of indicator. By selecting the Stepwise method for each block, the variable(s) that contribute most to the explanation of the dependent variable are determined. Using this second method, the final model is more likely to contain variables out of different indicator categories. A final approach first determines the most important safety performance indicators (using the Stepwise method). These indicators are included in the model in a first block (using the Enter method). Subsequently, the indicators of the other categories are stepwise included in a second block and therefore increase the explanatory power of the model.

The resulting models are tested on the error term assumptions of homoscedasticity, no autocorrelation and normality. In addition, the statistical significance of each model and the regression coefficients, the correlation between variables and the degree of multicollinearity are studied in order to guarantee acceptable models. Finally, the models that meet these different preconditions are considered (see below) and the explanatory power of the model is assessed.

3.1.2. Results

Three significant regression models (see Table 1) remain meeting the assumptions of homoscedasticity, no autocorrelation, normality and no (high) multicollinearity. Next, the explanatory power of each model is determined by using the adjusted determination coefficient ($R^{2'}$) that takes into account the number of variables used and the sample size (Hair et al., 1998).

Table 1 : Regression models

Model 1:	
FATAL = 0.183 – 0.774 E1 – 0.440 ALC + ε	$R^{2'}$: 78 %
Model 2:	
FATAL = 0.046 – 0.378 E1 – 0.376 PS + 0.413 D4 + 0.284 T1 – 0.264 T12 + ε	$R^{2'}$: 83.9 %
Model 3:	
FATAL = 0.109 – 0.336 PS – 0.369 VEH + 0.535 D4 + ε	$R^{2'}$: 71 %
<u>Legend:</u>	
• **FATAL**: # road fatalities / million inhabitants • **E1**: Gross domestic product • **ALC**: % drivers < alcohol limit • **PS** : % seat belt use in front of cars or vans	• **D4** : Avg. # persons / family • **T1**: # passenger cars /1000 inhabitants • **T12**: Avg. # kms driven per driver • **VEH**: % cars < 6 years

Before interpreting these models, it is important to note that they are not generalizable because of the small sample and low variable-observation ratio in this research. Nevertheless, these models are valuable because they offer information on the importance of each indicator and the nature of the relationship with the number of road fatalities per million inhabitants. This is achieved by interpreting the direction and size of the regression coefficients. We now concentrate on the different indicators in the models and their coefficients.

Explanatory road safety (performance) indicators are investigated first. For each of the seven risk domains defined in the European project SafetyNet (2005) (alcohol & drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management), one performance indicator was formulated. The results show that the risk dimensions 'alcohol & drugs', 'protective systems' and 'vehicle' have the most explanatory power with respect to the number of road fatalities per million inhabitants. These indicators have a negative coefficient, which means that a high score on these performance indicators, leads to less road fatalities per million inhabitants. An improvement in the share of drivers respecting the legal alcohol limit, a higher seat belt rate in front of cars or vans and a higher share of new (<6 years) cars will enhance the final road safety output level in a country. Apart from these road safety performance indicators, a number of background and context characteristics are of importance, namely 'the gross domestic product', 'the average number of persons within a family', 'the number of passenger cars per 1000 inhabitants' and 'the average number of kilometers travelled by a driver' (see Table 1). These features affect the final road safety level as well. Larger households and more cars per inhabitants have a negative impact on the road safety level. A good economic situation (expressed as a high GDP) and a high average amount of kilometres travelled by drivers lower the number of fatalities per million inhabitants. In contrast to the safety performance indicators, aspects related to the background and context are less controllable. Moreover, taking action with respect to safety performance (e.g. more controls) is interesting as it will have a direct effect on the number of road fatalities per million inhabitants.

3.2. Describing and comparing countries

Next, two techniques for describing and comparing countries are discussed. Firstly, clustering which enables us to discover similarities and differences between European countries, is discussed. Secondly, two summary footprints are illustrated for Belgium.

3.2.1. Clustering

We start by briefly discussing some theoretical and methodological considerations concerning clustering. More information can be found in Hermans et al. (2009). Afterwards, the results of the clustering based on safety performance indicators and safety output indicators are presented.

a) Theoretical background and methodology
Kaufman & Rousseeuw (2005) describe cluster analysis as "the art of finding groups in data". Clustering is putting objects (countries in this case) in different homogenous groups (subsets or clusters) with similar features, by making use of algorithms. That way, a simplified structure is created allowing to discover similarities and differences between groups of objects (Hair et al., 1998). In this research, we group the 21 countries in our data set according to each of the seven categories of indicators (see Figure 1). The different cluster results are compared to assess whether countries that are grouped in the same cluster for a certain set of indicators also belong together regarding another set of indicators.

To benefit from the advantages of both hierarchical and non-hierarchical clustering, these methods are combined (Hair et al., 1998). First, hierarchical clustering is performed, using the Ward algorithm, to determine the ideal number of groups (given 21 countries to cluster we are particularly interested in the range of three to six clusters). The agglomeration scheme generated in the output, which gives information on the objects being combined at each stage of the process and the corresponding agglomeration coefficient, is analyzed (Hair, 1998; Malhotra & Birks, 2003). By calculating the largest percentage change in the agglomeration coefficient (when going from n to n-1 groups) the optimal number of groups can be found (Hair et al., 1998). Then, non-hierarchical clustering (the k-means algorithm) is performed, using this number of groups as input. Due to missing values, the indicator set used in each cluster analysis is limited to guarantee that enough countries are incorporated in the analysis. After performing the cluster analysis, the grouping is studied to check for possible outliers significantly affecting the clustering results. In case a cluster consists of only one country that has very specific indicator values, the analysis can be redone and the results compared to decide on the final cluster result. Finally, the grouping of countries is visualized on a map. In the interpretation of the results, special attention is paid to the clustering of Belgium as well as the best performing European countries.

b) Results

In this paper, we present the results of grouping countries based on road safety output and road safety performance indicators. Concerning the background and context indicators only the final results are taken into consideration to draw conclusions. More detailed information concerning the grouping of background and context indicators can be found in Hermans et al. (2009).

Clustering based on seven road safety performance indicators (SPI's)

Countries are grouped with respect to safety performance by applying the methodology described above. The first results indicated that the cluster structure is influenced by two countries namely Greece and Cyprus, both having indicator values that differ significantly from the ones of the other European countries. Because the grouping of the other countries is influenced by these outliers, Greece and Cyprus will no longer be included in the analysis. Grouping the remaining 19 countries according to their value on the seven performance indicators, gives the following clustering result:

Figure 2 : Four clusters based on safety performance indicators

It can be seen that the Czech Republic, Hungary, Poland and Estonia are grouped together in one cluster. A second cluster consists of the countries Belgium, Spain, Italy and Portugal. In other words, concerning safety performance Belgium shows some similarities with these three South-European countries. Switzerland, Germany, France and the Netherlands are grouped in a third cluster. The other countries (Austria, Denmark, Finland, Ireland, Sweden, Slovenia and United Kingdom) form a fourth group.

Cluster one - consisting of the Czech Republic, Hungary, Poland and Estonia - distinguishes itself positively from the other clusters for the risk domain 'alcohol & drugs'. This implies a high share of drivers stating to respect the prevalent alcohol limit in their country. On the contrary, these countries are characterized by relatively less recent vehicles and a low seat belt wearing rate in front of cars and vans. Studying the second cluster (Belgium, Spain, Portugal and Italy) we see average scores on the risk domains 'roads', 'trauma management' and 'vehicle' and low scores on the risk domains 'alcohol & drugs' and 'speed'. Furthermore, the seat belt use in front of cars and vans is relatively low comparing to other groups. The countries in cluster three (Switzerland, Germany, France and the Netherlands) distinguish themselves due to their high performance on the risk domains 'trauma management', 'roads' and 'protective systems'. Countries in cluster four (Austria, Denmark, Finland, Ireland, Sweden, Slovenia and United Kingdom) perform well on all risk domains except 'trauma management' and 'roads'.

A close look at the safety performance indicator values of Belgium and the SUN-countries reveals that despite the fact that Belgium scores almost as high as the Netherlands for the risk domain 'roads', the Netherlands scores significantly better on the other risk domains. Therefore, Belgium is not grouped together with the Netherlands. Two of the SUN-countries, United Kingdom and Sweden, are clustered because of their similarly good scores on the risk domains 'alcohol & drugs', 'speed', 'protective systems', 'vehicle' and 'daytime running lights'. The Netherlands belongs to another cluster mainly because of its high motorway density ('roads').

Clustering based on five road safety output indicators

Next, the similarities and dissimilarities between countries in terms of road safety output indicators are studied. In this paper, we present the clustering results based on the following five indicators: fatalities per million inhabitants, fatalities per 100,000 cars, fatalities per 100 million passenger kilometers, injury accident per 100,000 inhabitants and the average yearly reduction in fatalities in percentages. Some values are missing for Switzerland and Cyprus, so in total 19 European countries are grouped in three clusters:

Figure 3 : Three clusters based on safety output indicators

Figure 3 shows that Belgium forms cluster together with Austria, Spain, Italy, Portugal and Slovenia. In a second cluster, the best performing European countries (Sweden, United Kingdom and the Netherlands) are grouped together with Denmark, Germany, Finland, France and Ireland. The worst performing countries regarding the five safety output indicators (the Czech Republic, Greece, Hungary, Poland and Estonia) constitute a third cluster.

The countries in the second cluster have the best road safety outcomes (or the lowest relative number of fatalities, etc). Germany, Belgium, Spain, Italy, Portugal and Slovenia (cluster one) take the second position within Europe and score average and have a similar average yearly reduction in fatalities as the best performing countries. Furthermore, the number of injury crashes per 100,000 inhabitants is relatively high for these six countries. The worst performing countries can be found in cluster three (the Czech Republic, Greece, Hungary, Poland and Estonia) having a high fatality rate yet an average number of injury crashes per 100,000 inhabitants. The latter might be due to a low registration level of (less serious) crashes.

The same cluster analyses as described above, are performed for the background and context indicators. In general, it can be concluded that Belgium shows some resemblances with neighboring country the Netherlands (among other things with respect to demography, casualty type and location of crashes) yet performs worse on a number of risk dimensions. Consequently, concerning road safety performance and road safety output Belgium does not belong to the same cluster as the best performing European countries but is grouped with a number of South-European countries. Moreover, it can be deduced that the SUN-countries are not only alike with respect to road safety output indicators but also have similarities with respect to road safety performance indicators and economical and transportation background.

3.2.2. Footprint

Next, the footprint methodology developed by Morsink et al. (2005) is briefly described and applied to the case of Belgium.

a) Theoretical background and methodology

Indicators can be structured in a road safety footprint; this schematic overview is useful for describing and comparing countries. The summary footprint presents the key characteristics of a country in a certain time period in the form of a table which consists of context and background indicators in the first column, road safety performance indicator in the second column and road safety output indicators in a last column. Because footprints give a quick and compact view of the road safety level in a country, they are a valuable instrument for policy makers (Morsink, et al., 2005).

In this research, we set up two summary footprints for Belgium. One footprint compares Belgium with the

European average ('general footprint') and the other footprint compares Belgium with the average of the SUN-countries ('detailed footprint'). To visualise the relative score on an indicator (compared to the European or SUN average) a colour (red, orange or green) is assigned to every indicator. Therefore, boundaries for the colouring are decided upon on the basis of minimum, maximum and average scores (see Hermans et al., 2009). Aspects on which Belgium scores more or less average are indicated by the orange colour. Important to notice is that a higher indicator value then the average value (of Europe or the SUN-countries) does not necessarily imply a green colour. For example, an above-average share of young people (15-24 years) within the population is disadvantageous from a road safety point of view - due to their lack of driving experience, risky behaviour, etc (Evans, 2004) - and consequently coloured red.

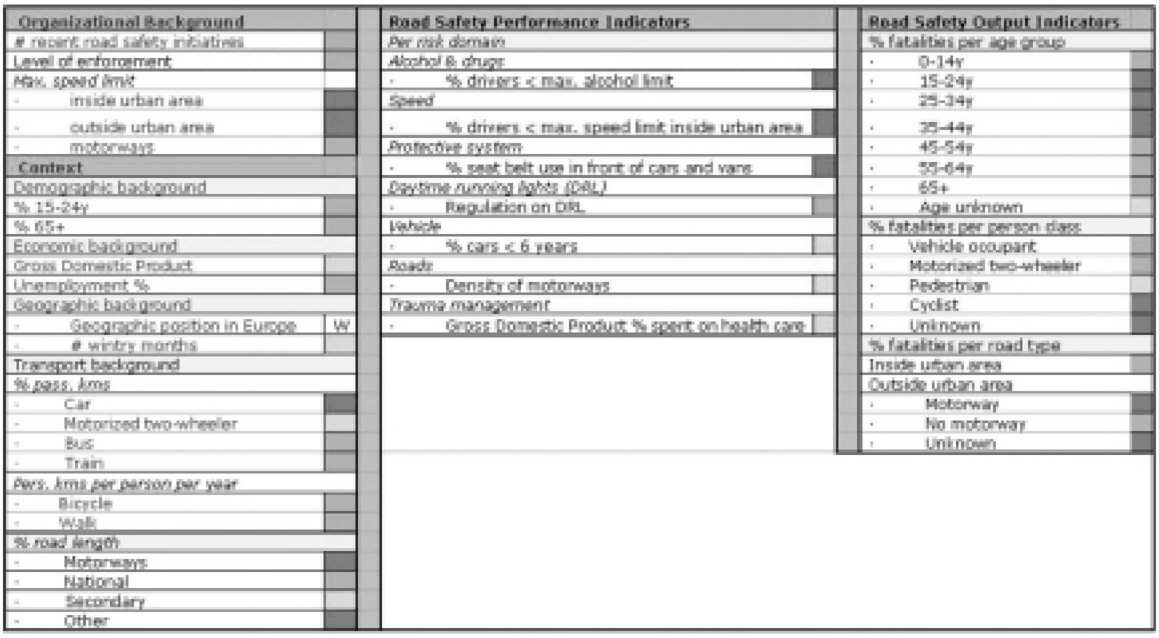
b) Results

We will now present the results of the two footprints. First, the general footprint, comparing Belgium to the average of the European countries in the data set, is discussed. We concentrate on the aspects on which Belgium scores significantly better and significantly worse than the European average.

General footprint

When we look at the organizational background and context part in Figure 4 it can be deduced that Belgium scores, compared to the European average, relatively good on the gross domestic product (above average), the number of wintry months (below average) and the percentage of passenger kilometers driven with a motorized two-wheeler (below average). Belgium scores significantly worse concerning the prevalent speed limit inside and outside the urban area (above average) and the number of passenger kilometers driven with cars (above average). In addition, the share of motorways in terms of road length is significantly lower and that of other roads significantly higher, both negative aspects from a road safety perspective.

Figure 4 : General footprint: Belgium - European average



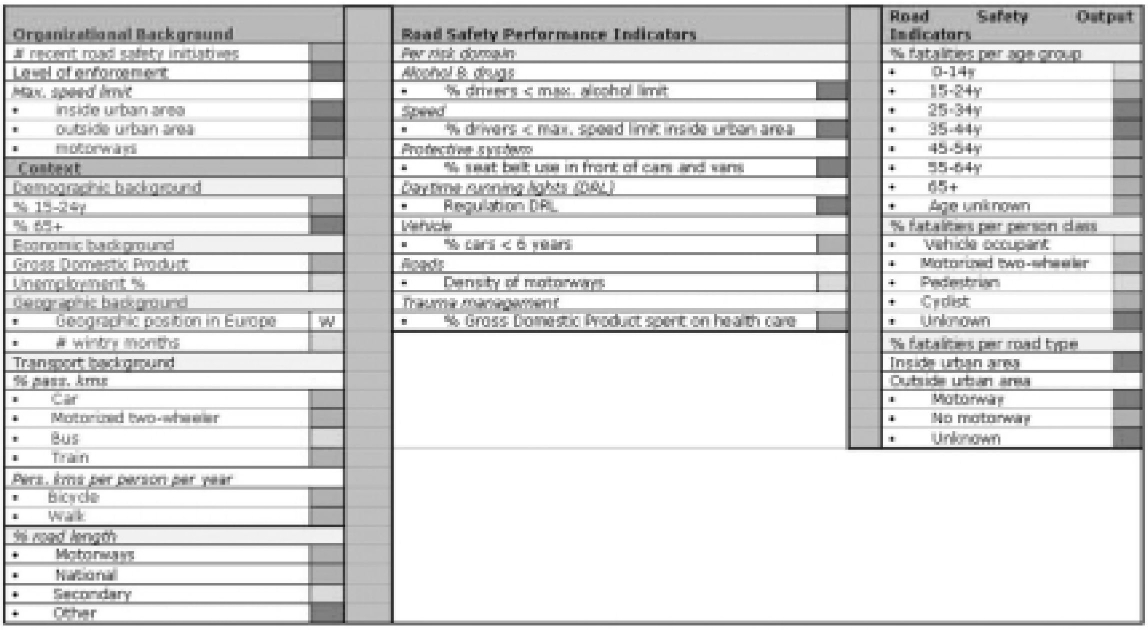
Concerning the SPI's, Belgium performs well on the risk domains 'vehicle', 'roads' and 'trauma management' yet underperforms on the behavioural risk domains 'alcohol & drugs', 'speed' and 'protective systems'. In other words, less drivers (state to) respect the legal alcohol limit and the speed limit inside urban areas and the seat belt wearing rate in front of cars and vans is lower, compared to the European average. Measures aimed at an improvement in these indicator values such as a higher level of enforcement will have a positive effect on the final road safety level.

Relatively more people within the age groups 15 until 44 years become a fatal victim of road crashes in Belgium. Taking into account the percentage of young people (15-24 years) in the population (which is comparable to the European average) based on the high share of fatalities among this age group, we can conclude that young persons clearly constitute an important risk group in Belgium and therefore deserve special attention. When we consider the distribution of fatalities per vehicle occupant, motorized two-wheeler, pedestrian, cyclist and unknown person class we find that the share of fatalities among car occupants approaches the European average, given the fact that relatively more passenger kilometers are travelled by car in Belgium. The share of fatalities among motorized two-wheelers also approaches the European average, but because a significantly lower part of the passenger kilometers are travelled with this vehicle type, motorcyclists and moped riders are another risk group. Furthermore, Belgium has fewer fatalities among pedestrians and more fatalities among cyclists. In Belgium, walking is relatively safer than cycling (taking into account the number of kilometers travelled per person within a year). Next, there are relatively more fatalities on motorways, despite the fact that the road length of motorways is small compared to the European average. The actual amount of travel on this road type would offer valuable information in this respect.

Detailed footprint

Next, the detailed footprint comparing Belgium with the SUN-countries is shown. In the discussion we focus on the main differences concerning safety performance and safety output indicators.

Figure 5: Detailed footprint Belgium - SUN-countries



Since we benchmark the Belgian (road safety) situation with respect to that of the best performing European countries there are fewer aspects on which Belgium performs better. The most important indicators for which it does are the risk domain ‘roads’, the share of fatalities between 0 and 14 years old and the share of fatalities among car occupants and pedestrians (given comparable passenger or person kilometres travelled by these modes). Especially the fact that Belgium performs poorly on most risk domains contributes to its worse road safety output. Remarkable is that young persons (aged between 15 and 24 years) also form a risk group in the best performing European countries (particularly in United Kingdom). The share of children among the fatalities is higher in the best performing countries than in Belgium. Compared to the best performing countries, the share of fatalities as vehicle occupant and pedestrian is lower in Belgium. However, the larger category of ‘unknown person class’ could explain these last differences. Finally, Belgium has more fatalities inside urban areas and on motorways.

Although this third analysis emphasized on Belgium, this footprint methodology can be applied to every country and as a result its best and worst characteristics compared to other countries can be identified. The summary footprints, in which Belgium is compared to the average European situation on the one hand and to the SUN-countries on the other, revealed that Belgium underperforms in many areas. The most important problem areas concerning safety performance are the low seat belt rate (in front of cars and vans) and the low performance on the risk dimensions ‘alcohol and drugs’ as well as ‘speed’. The road safety output shows among other things the relatively high share of fatalities between 15 and 44 years and the high share of fatalities on motorways.

4. Conclusion

In this paper, different techniques for describing, explaining and comparing the road safety situation in a country are presented and illustrated. These - regression, cluster and footprint - techniques can be applied to gain insight into various data sets. In this research, road safety related indicators belonging to seven categories (safety output indicators, safety performance indicators, geographic indicators, demographic indicators, economic indicators, transport indicators and mobility policy indicators) were formulated and data were collected for 21 European countries.

When explaining the number of road fatalities per million inhabitants the following indicators appeared to have the most explanatory power: ‘alcohol and drugs’ (% drivers < alcohol limit), ‘protective systems’ (% seat belt use in front of cars and vans), ‘vehicle’ (% cars < 6 years), ‘the average number of persons within a family’, ‘the gross domestic product’, ‘the average number of kilometers traveled by a driver’ and ‘the number of passenger cars per 1000 inhabitants’. Taking action on the risk dimensions is interesting as it will have a direct effect on the number of road fatalities per million inhabitants. Next, countries were grouped in order to identify similarly performing countries. The results showed that Belgium is often grouped with South-European countries. The SUN-countries show apart from their similar road safety level some resemblances concerning the economic and transport background. Finally, the footprints for Belgium described and compared it with the European average and the average of the SUN-countries. They revealed important problem areas and are therefore a valuable instrument for policy makers.

In the future, the indicator set can be extended with other relevant indicators and data can be gathered for more countries. The availability and quality of road safety data always plays an important role. Further research could focus on the imputation of data.

5. References

Al Haji, G. (2005). Towards a road safety development index. Ph.D. thesis, Linkopings universitet.

European Transport Safety Council (2001). Transport Safety Performance Indicators. ETSC.

European Union Road Federation (2006). European road statistics 2006. ERF.

Evans, L. (2004). Traffic Safety. Bloomfield, Michigan: Science Serving Society.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. (1998). Multivariate data analysis. New Jersey: Prentice Hall, Inc.

Hermans, E. (2009). A methodology for developing a composite road safety performance index for cross-country comparison. Ph.D. thesis, Hasselt University.

Hermans, E., Wilmots, B. & Brijs, T. (2009) (in Dutch). Europese vergelijking op basis van verkeersveiligheidsindicatoren: Beschrijving, verklaring, groepering. Steunpunt Mobiliteit en Openbare Werken - Spoor Verkeersveiligheid.

Kaufman, L. & Rousseeuw, P.J. (2005). Finding groups in data: An introduction to cluster analysis. New York: Wiley-Interscience.

Malhotra, N.K. & Birks, D.F. (2003). Marketing research: An applied approach. London: Prentice Hall, Inc.

Morsink, P., Oppe, S., Reurings, M. & Wegman, F. (2005). SUNFlower+6: Development and application of a footprint methodology for the SUNflower+6 countries. SWOV.

SafetyNet (2005). State of the art report on road safety performance indicators: D3.1.

SafetyNet (2008). Traffic Safety Basic Facts 2008: Main Figures.

Sartre 3 consortium. European drivers and road risk. INRETS.

Sivak, M. & Tsimhoni, O. (2008). Improving traffic safety: Conceptual considerations for successful action. Journal of Safety Research 39, pp. 453 - 457.

World Health Organization (2009). Global status report on road safety: Time for action. WHO.