DOCTORAATSPROEFSCHRIFT

2009 | Interfacultair Instituut Verkeerskunde

A methodology for developing a composite road safety performance index for cross-country comparison

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Verkeerskunde, te verdedigen door:

Elke HERMANS

Promotor: prof. dr. Geert Wets

universiteit hasselt

INSTITUUT VOOR MOBILITEIT

DOCTORAATSPROEFSCHRIFT

2009 | Interfacultair Instituut Verkeerskunde

A methodology for developing a composite road safety performance index for cross-country comparison

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Verkeerskunde, te verdedigen door:

Elke HERMANS

Promotor: prof. dr. Geert Wets

universiteit hasselt
INSTITUUT VOOR MOBILITEIT

D/2009/2451/11

"The formulation of a problem is often
more essential than its solution
which may be merely a matter of
mathematical or experimental skill..."
(Einstein & Infield, 1938)

Preface

In the past years, I have performed research in the area of road safety. In this dissertation, the knowledge acquired during that period is compiled. This manuscript aims at making a valuable contribution to the topic of a composite road safety performance index by developing a scientifically sound and appropriate index methodology.

This research was partly made possible by the Policy Research Centre Mobility and Public Works, track Traffic Safety. With respect to this research in general and this dissertation in particular, some expressions of gratitude are to be called for. In the first place, I am indebted to the members of my Ph.D. commission, prof. dr. Geert Wets, prof. dr. Koen Vanhoof, prof. dr. Tom Brijs en dr. Filip Van den Bossche. They gave me the opportunity to carry out challenging scientific research and guided me with their insight throughout this process. In the second place, I am grateful to the external members of the jury, prof. dr. Shalom Hakkert, dr. ir. Martijn Vis and prof. dr. Da Ruan for their constructive comments on this subject.

In addition to the professional contribution of these persons, I show my gratitude to my colleagues at IMOB for creating an excellent working environment. The many joyful moments shared together as well as their support gave me additional strength to bring this doctoral research to a favourable end.

Evidently, I thank my parents, in-laws, family and friends for showing interest in my work, offering me distraction and being understanding especially during the final phase of this dissertation.

Finally, I owe a large debt of gratitude to my friend Jeroen, for his patience and continuous support throughout the past years.

Elke Hermans
April 27th, 2009

Executive Summary

In this research, countries are compared in terms of road safety. Hence, the complex and multidisciplinary concept of road safety is translated into quantifiable indicators. Up to now, country comparisons were mainly based on registered accident data (i.e., final outcome information). However, we want to gain insight into the main underlying road safety risk aspects. That way, appropriate measures able to tackle the main risk aspects can be selected before they result in an increased number of accidents and casualties. For that reason, road safety performance indicators situated at the intermediate outcome level are studied here. In particular, starting from seven main risk domains identified at the European level, indicators related to alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management are considered.

In order to compare countries on their overall road safety performance we focus on the combination of indicators into a composite index. One of the advantages of an index over a set of individual indicators is the ability to present a multitude of information in one overall score thereby facilitating benchmarking and communication. Contrary to other research fields, very limited research effort has been put in the creation of a road safety index. The main objective of this manuscript is to develop a scientifically sound and appropriate methodology for the creation of a composite road safety performance index that can be used for the monitoring and comparison of road safety performance among countries. The different methodological aspects involved in the index construction process are described in terms of theory and illustrated using a real road safety indicator data set. Based on the index score – taking the performance on several road safety risk aspects into account rather than solely the number of road fatalities – valuable insight is gained into the relative road safety performance of a set of European countries.

After introducing the concept of indicators and the road safety problem, this manuscript outlines the different methodological steps involved in the index process and successively discusses each step in detail. Below, the main steps are briefly described.

1. Selecting appropriate indicators to combine in an index: for each risk domain, possible indicators were listed from literature and subsequently evaluated taking several selection criteria into account. Each indicator was judged on relevance, measurability, interpretability, specificity and sensitivity thereby resulting in the identification of ideal or best needed indicators. However, to illustrate the index methodology on a real data set, three additional criteria were taken into account, i.e., data availability, reliability and comparability. The created index was composed of so-called best available indicators.

2. Collecting indicator data: various data sources were consulted in order to collect data for the indicators selected in the previous step. A data set containing road safety performance indicator values for a large set of countries, referring to a particular period or (preferably the most recent) year, was aimed at. The degree of maturity of performance indicators as well as the level of data availability and quality varied between the risk domains. In the end, 2003 data for six indicators reflecting essential and best available information to track road safety performance were obtained for 21 European countries. The direction of the indicators was geared to one another with higher values (i.e., a better road safety performance) to aim at.

3. Performing data analyses: analyses were performed on the collected indicator data set in order to gain more insight into each indicator separately (univariate analysis) as well as into the structure and interrelationships of the whole indicator set (multivariate analysis). Summary statistics on each indicator were presented; normalisation techniques to make the data comparable over the indicators were discussed; and theoretical considerations on the issue of missing values were given. Based on the results of the multivariate analysis the appropriateness of the indicator set for combination in an index was assessed (i.e., it has a sufficient level of correlation, internal consistency, clustering and explanatory power).

4. Assigning a weight to each indicator: the theory behind five common weighting methods – factor analysis, analytic hierarchy

process, budget allocation, data envelopment analysis and equal weighting – was described as well as their advantages, disadvantages and requirements. Weights obtained by means of factor analysis are based on correlations; analytic hierarchy process and budget allocation weights result from expert's opinions questioned in a different way (i.e., relative contribution per pair of indicators respectively budget distribution); data envelopment analysis (DEA) weights are deduced from a constrained optimization model; and equal weighting assigns a weight of 1/6 to every indicator. Each method was applied, the resulting index scores computed and the final countries' rankings compared. A qualitative assessment of the five methods (in terms of additional insights) as well as a quantitative one (based on the correlation coefficient between each index ranking and the final outcome ranking) disclosed that the data envelopment analysis method is the most promising weighting method of the five in the road safety index context.

5. Aggregating indicators: the mathematical formula for combining the indicators into an index needed to be selected. In addition to the often applied linear and geometric aggregation, the aggregation research field was taken as starting point. In general, the class of averaging operators – in which the index score is bounded by the lowest and highest indicator value – and in particular, the ordered weighted averaging (OWA) operators – assigning weights to magnitudes of performances – appeared to be the most useful aggregation operators for the road safety index case. Special types of OWA operators (e.g., the maximum operator) were discussed and different OWA weighting vectors were judged. The resulting vector reflected a rather intolerant aggregation idea in which all performances were considered yet worst performances were emphasized.

6. Testing the robustness of the index: given the multistage index process, the index developer is left with a number of methodological choices that might influence the final countries' ranking. Performing uncertainty and sensitivity analyses is an essential part of the index process as it shows how robust the index is and which extra information would imply more

robustness. The impact of the methodological choices made during the index construction process – with respect to indicator selection, normalisation technique, weighting method, expert selection and way of aggregating – on the absolute average shift in countries' ranking based on the index scores compared to a reference ranking was quantified. Two reference rankings, i.e., the default index ranking and the final outcome ranking were considered. Moreover, the sensitivity indexes revealed the large impact of the weighting method and the indicator selection. Hence, they require careful evaluation and justification.

7. Computing, evaluating and visualising final index scores: taking all previously acquired information into account a final index methodology was presented. More specifically, the most promising weighting method – data envelopment analysis – was further elaborated in terms of aggregation. For each country, the most optimal index score under the imposed restrictions was determined. This index score was at the same time an acceptable solution due to the incorporation of weight boundaries regarding the share of each indicator in the index score (deduced from budget allocating experts) and the emphasizing of bad performances (by means of an orness restriction). Applying the suggested index methodology on the best available data set (and keeping the limitations in mind) three groups of countries could be distinguished. The Netherlands, Germany, Switzerland, Denmark, Slovenia, United Kingdom and France obtained a high road safety performance index score; Ireland, Sweden, Austria, Belgium, Portugal, Finland and Spain had an average index score; and Czech Republic, Greece, Italy, Poland, Hungary, Cyprus and Estonia were characterized by a low index score. Furthermore, the relationship between the constructed index and other related indicators and indexes was assessed. A high degree of consistency between the countries' ranking based on the road safety performance index on the one hand and the final outcome, human development index and corruption perceptions index ranking on the other hand was found. The index process was closed with a visualisation of the final results.

In general, it can be concluded that insight was gained in the index construction process. Each step in this process has been thoroughly examined. From literature, relevant options for the road safety index case were selected and profoundly evaluated. Supported by qualitative and quantitative assessment a scientifically sound and appropriate index methodology was proposed for the combination of best available road safety performance indicators in a composite index. The most optimal index score for each country was computed thereby respecting imposed restrictions concerning the relative importance of an indicator and the relative magnitude of a performance.

During this research, the link between the road safety performance index (the intermediate outcome layer) and the number of road fatalities per million inhabitants (the final outcome layer) has been kept in mind. The level of agreement between the two demonstrates that research on the intermediate outcome level is promising and relevant because the road safety performance index and its indicators can act as warning signals and enable earlier, goal-oriented action. In addition to the discussion of the index process, one chapter was devoted to the determination of a road safety score that represents both intermediate outcome indicators and final outcome indicators. By developing an input-output data envelopment analysis model, countries could be ranked based on a score capturing different types of road safety indicators. By taking the characteristics of each country into account, country-specific road safety enhancing recommendations in terms of benchmarks and priorities could be made.

To conclude, a composite road safety performance index is a valuable communication, benchmarking, monitoring and policy supporting tool for comparing road safety performance across countries. A scientifically sound and appropriate methodology as well as good road safety performance indicators are a prerequisite in this respect. The future gain in importance of concepts like road safety performance indicator and road safety performance index will give rise to new data opportunities whereas the research described in this manuscript may make a valuable contribution to future road safety index development.

Table of contents

List of figures

List of tables

Chapter 1 Introduction

Road safety is a topic that is correctly receiving a lot of attention lately. Given the high number of casualties and the corresponding suffering and costs, measures are needed in order to reduce the number of road casualties. Better insight into the road safety situation of a particular subject (in this study we focus on the level of countries) can be gained by studying the available data and comparing them to the data of other subjects. Nowadays, this country comparison in terms of road safety is mainly based on registered accident data. For example, the number of injury accidents and the number of casualties (divided into fatalities, serious casualties and slight casualties) per capita are computed and evaluated. The relative position of each country can then be assessed. However, these accident related figures are unable to indicate on which aspects of road safety an underperforming country should focus in order to improve its road safety level. Therefore, countries should be compared on a more detailed level as well.

More specifically, the main underlying risk factors of road safety need to be determined. Each risk factor (e.g., alcohol and drugs) should be represented by appropriate safety performance indicators (e.g., the share of road users with a blood alcohol content above the legal limit). The indicator values can be compared across countries thereby resulting in the identification of the main problem areas in a particular country. Appropriate measures can then be selected able to tackle the main risk aspects before they result in accidents and casualties.

Countries could be compared on each indicator individually. However, given the high number of relevant road safety performance indicators the creation of a composite road safety performance index – which is a combination of road safety performance indicators – is valuable. Contrary to a set of individual indicators an index presents the overall road safety picture by capturing a multitude of risk information in one index score. The creation of an index score for each country enables easy comparison across countries. In the recently published SUNflowerNext report (Wegman et al., 2008) a composite index is characterized by the words 'simplification', 'quantification' and 'communication'.

This manuscript deals with the combination of road safety performance indicators into one index. The major objective of this research is **to develop a scientifically sound and appropriate methodology for the creation of a composite road safety performance index that can be used for the monitoring and comparison of road safety performance among countries**. The different methodological aspects involved in the index process are described in terms of theory and applied to a real road safety indicator data set in order to gain insight into the relative road safety performance of various European countries.

This first chapter starts with a description of some basic concepts (Section 1.1) followed by the purposes of indicators (Section 1.2). Section 1.3 deals with the overall planning process. Next, the road safety problem (Section 1.4) and road safety indicator frameworks (Section 1.5) are described. In Section 1.6, the type of indicators focused on in this research – road safety performance indicators – are discussed and the chapter closes with a description of the structure of this manuscript (Section 1.7).

1.1 Basic concepts

In this section, some basic concepts used throughout this manuscript are defined. The concepts refer to the indicator literature in general (European Commission, 2005; Litman, 2005; Nardo et al., 2005b), road safety in general (Al Haji, 2005; Morsink et al., 2007) and road safety indicators (European Transport Safety Council, 2001).

- Indicator: a measure that summarizes an aspect of a phenomenon in an understandable and relevant way
- Indicator data: values used for indicators
- Indicator framework: conceptual model for structuring indicators and their interrelationships
- Indicator set: a group of indicators selected to measure an overall phenomenon

- Best needed indicator: ideal indicator[1] having one or a few shortcomings such as no data available, no reliable data or no comparable/coherent data
- Best available indicator: best possible[2] indicator for which the available data are of acceptable quality
- Composite index[3]: a combination of an indicator set
- Normalisation: data transformation in which the indicators (e.g., expressed in different measurement units) are rendered comparable
- Aggregation: mathematical formula for combining an indicator set into an index
- Robustness: stability in the output given small changes in the input
- Target: a specific, measurable, achievable, relevant and timed objective
- Benchmark: a subject with a higher success ratio compared to other subjects; this best-in-class can be seen as an example for ameliorating performance
- Casualties: the number of (fatally, seriously or slightly) injured persons from a road accident
- Mortality rate: the number of road fatalities in a country divided by the number of inhabitants living in that country (per 100,000 inhabitants)
- Fatality rate: the number of road fatalities in a country divided by the number of motorized vehicles (per 10,000 vehicles)

[1] Among the evaluated indicators, this indicator scores best in terms of relevance, measurability, interpretability, specificity and sensitivity.

[2] Among the evaluated indicators, this indicator scores best in terms of relevance, measurability, interpretability, data availability, reliability, comparability, specificity and sensitivity.

[3] The term 'index' refers to the same in this manuscript.

- Fatality risk: the number of road fatalities in a country divided by the number of motorized vehicle kilometres (per 100 million vehicle kilometres)
- Safety performance: the level of transport safety. A reduction in the number of accidents or the number of casualties corresponds to an amelioration of the safety performance
- Safety performance indicator (SPI): any measurement that is causally related to traffic accidents or casualties, used in addition to a count of accidents or casualties in order to indicate safety performance or better understand the process that leads to accidents.

1.2 Purpose of indicators

In recent years, there has been an explosion of interest in indicators in several domains (such as environment (e.g., Hens et al., 2005), economy (e.g., Kaminsky et al., 1998), education (e.g., Studiedienst van de Vlaamse Regering, 2006) and innovation (Kleinknecht et al., 2002)). This reflects growing recognition of the important role indicators can play as a tool for enhancing the quality of decision making. Indicators express scientific knowledge in a form that supports decision makers to take better informed and more appropriate choices. From literature (Al Haji, 2005; Hens et al., 2005; Litman, 2005; Nardo et al., 2005b; Organisation for Economic Co-operation and Development, 2001; Salzman, 2003; Van Reeth and Vanongeval, 2005) it appears that indicators can be used in a number of ways, such as:

- measuring relative performance/benchmarking: indicators are measures derived from a series of observed facts that can reveal relative positions in a given area. Best-in-class and superior performance can be established.
- drawing attention to particular issues: indicators are suited for communication purposes, such as informing policymakers or activating and stimulating the public alertness.
- identifying trends: in case an indicator is measured at regular intervals, the directions of change over time and across different subjects can be pointed out.

- predicting problems: indicators can serve as warning signal for policymakers and are important guidelines for governments and authorities.
- assessing the impact of policy measures: indicators can be used for evaluating intended output and policy effects, enabling to judge several options.
- setting targets and priorities: based on former indicator values and values from other subjects, targets can be set. A comparison between different indicators may reveal which aspects need (more) urgent action.
- evaluating progress towards targets: in case of indicator measurements at regular moments in time the progress towards stated targets can be monitored closely and the achievement estimated. At certain time points, re-evaluation of goals or remedial action might be appropriate.
- presenting in a comprehensible way: indicators can present a large amount of information in a clear way. They are means for visualising the current situation. That way, problems become more concrete and subject to discussion. Several ways of presenting indicator information exist and will be discussed later.

The idea of using indicators for the continuous monitoring and analysis of processes exists for decades. Modern use of performance measures rose out of the Deming total quality management movements of the 1950s in Japan. The principles rely on developing goals that can be related to measurable results, monitoring those results and assessing strategies to improve performance (National Cooperative Highway Research Program, 2003).

1.3 The planning process

Indicators are useful tools in terms of monitoring, evaluation and communication. The identification of indicators is one step in a broader planning process. In general, a safety planning and decision making process is characterized by the following phases (Federal Highway Administration, 2004): understanding the problem; establishing institutional leadership, responsibility and accountability; defining desired outcomes; identifying indicators; comparing with other experiences; developing and implementing a systematic safety data collection and analysis process; developing a safety plan and integrating it into decision making; and monitoring the effectiveness of implemented strategies and actions. Each step is briefly described below in relation to Belgium/Flanders.

1. understand the problem: in order to come up with strategies and investments for enhancing the level of road safety, the problem should be understood first. In order to gain insight into the size of the road safety problem and its contributing factors, accident data can be analyzed. At the aggregated level, the number of accidents or casualties can be plotted over time thereby revealing the overall trend. Looking at the accident causes and performing analyses at the disaggregated level (see e.g., Vlaams Ministerie van Mobiliteit en Openbare Werken, 2007) help in explaining these general trends and identifying the main problem areas or so-called risk domains.

2. establish institutional leadership, responsibility and accountability: attention should be paid to a clear division of responsibility, coherent sets of competences and subsidiarity. Arrangements between the authorities responsible for taking measures are needed as several levels (i.e., European, federal, regional and local) are involved. Apart from the ministers of mobility, the federal government service of mobility and transport and the Flemish department of mobility and public works there exist several advisory bodies (e.g., Mobiliteitsraad van Vlaanderen, Vlaams Forum Verkeersveiligheid), a number of institutes (e.g., Belgisch Instituut voor de Verkeersveiligheid, Steunpunt Mobiliteit en Openbare Werken – Spoor Verkeersveiligheid), pressure groups (e.g., Federatie van de Auto en Tweewielerindustrie, Fietsersbond), etc.

3. define desired outcomes: in case a target is set, there is a higher involvement and tendency to achieve it. Prediction models are often used in this respect as they offer interesting information for target setting. More specifically, the number of accidents and casualties in the (near) future can be predicted by means of time

series analysis (see e.g., Hermans et al., 2006c). The evolution in the past and other factors (such as planned interventions) can be incorporated in these kinds of models. Desired outcomes might be aggregated, e.g., maximum 500 road fatalities on Belgian roads by 2015 (Federale Commissie voor de Verkeersveiligheid, 2007) as well as disaggregated, i.e., related to specific user groups or focussed on particular causes of accidents, e.g., maximum 55 fatalities and fatally injured persons per 1 million persons under 26 years by 2010 (Ministerie van de Vlaamse Gemeenschap, 2001).

4. identify indicators: once a target has been established, the next step is to identify relevant indicators. Since 2004, the Belgian ministry of transportation issues a barometer on road safety at regular intervals. On the basis of police reports, the monthly evolution in registered accidents and fatalities on the spot is shown as well as statistics on casualties by region and road type. Recently, a barometer for accidents involving heavy good vehicles and accidents during the weekend has been set up. Moreover, performance indicators quantifying the risk domains leading to accidents and casualties (e.g., drinking and driving) can be formulated. In Chapter 3, the selection of safety performance indicators will be extensively discussed.

5. compare with other experiences: indicator data can be compared across regions or countries. The comparison process indicates the good and bad aspects of road safety in a country and reveals a benchmark that can be taken as an example.

6. develop and implement a systematic safety data collection and analysis process: a key to the success of national or regional road safety programs is the existence of a data collection and analysis system that provides frequent information on the safety level of the transportation system. This information is used to monitor progress towards targets, identify topics or areas where further action is necessary, educate officials and the public on the importance of the topic and evaluate the effectiveness of implemented projects and strategies. Nowadays, large amounts of information are available, for example related to the location

of accidents, weather characteristics, travel information, etc. Furthermore, collecting data on safety performance indicators is valuable for monitoring changes in the different road safety risk domains.

7. develop a safety plan and integrate it into decision making: the next step is to develop a road safety action plan outlining the road safety problems, challenges being faced, targets established, actions considered as well as institutional responsibilities for carrying out the plan. The key to a successful strategy is to find maximal safety at minimal cost; and to choose the best means of achieving it (Land Transport Safety Authority, 2000). Different measures can be fairly compared in terms of a cost-benefit analysis (see e.g., Hakkert and Wesemann, 2005). The safety plan usually consists of a range of strategies, actions and projects. For Flanders, seven main categories of measures have been stated (Vlaams Ministerie van Mobiliteit en Openbare Werken, 2007). Identifying the most appropriate combination of actions comes from both a technical process that identifies the likely effectiveness of different strategies and a political/public assessment of what might be feasible.
8. monitor effectiveness of implemented strategies and actions: determining the effectiveness of implemented strategies and actions is the eighth step. Monitoring outcomes is needed to see whether or not the forecast will be met (Land Transport Safety Authority, 2000).

This planning process needs to be revised from time to time. In the short run, a revision of the target or some extra measures might be required. In the long run, a careful review of the strategy including the assumptions and reasoning is needed as unforeseen circumstances (e.g., faster population or mobility growth, new technological developments) may push the strategy off course (National Road Safety Committee, 2000).

The identification of performance indicators is an essential yet largely uncovered issue at this moment. Therefore, Chapter 3 is devoted to this. First, the main risk domains for which safety performance indicators will be developed need to be identified. Therefore, the next section describes the road safety problem and its major determinants.

1.4 The road safety problem

1.4.1 Description

Road safety is an important topic as the price paid for mobility is too high. Worldwide, an estimated 1.2 million people are killed in road accidents each year and as many as 50 million are injured (World Health Organization, 2004). This means that every day around the world, more than 3,000 people die from road traffic injury. The economic cost of road accidents and casualties is estimated at 1 to 2% of the gross national product (World Health Organization, 2004).

In the European Union, every year more than 40,000 persons are killed and more than 1.7 million injured (European Commission, 2006a). Road traffic injuries cause physical, psychological, material and economic costs. The estimated annual cost of road traffic injury in European Union countries exceeds €180 billion (World Health Organization, 2004). Compared to other activities, the chance of dying in road traffic is per hour 40 times higher than at work and 12 times higher than during activities at home (Ministerie van de Vlaamse Gemeenschap, 2001). One person in three will be involved in a road accident with casualties at some point in his life (European Commission, 2001).

A closer look at the Belgian accident figures reveals that in 2007 the number of fatalities still exceeds 1,000 and the number of injured persons is almost 67,000 (http://www.statbel.fgov.be/downloads/accidents_dossier_2007_nl.xls). Compared to other European countries – the EU-27 average being 87 road fatalities per million inhabitants in 2006 – and particularly in relation to its neighbouring countries, Belgium scores relatively bad. In Figure 1, the number of road fatalities per million inhabitants in 2006 is given for 27 European countries (European Union Road Federation, 2008) and Flanders (FL) (http://www.statbel.fgov.be).

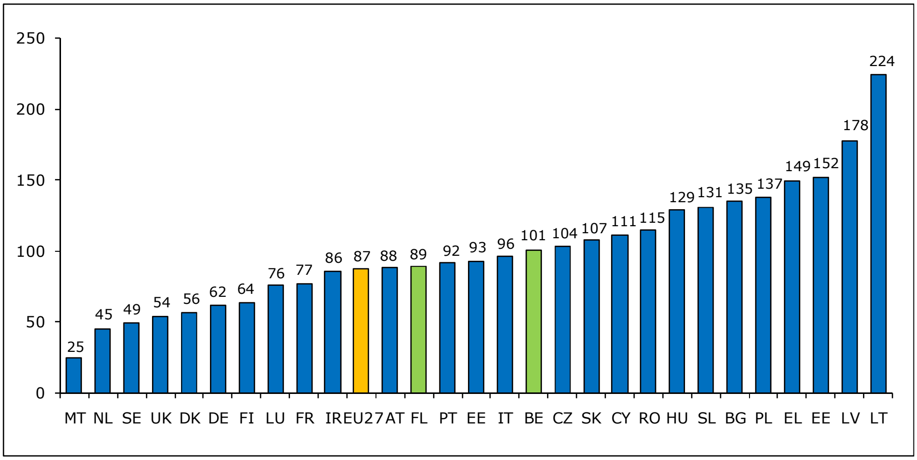


Figure 1: Road fatalities per million inhabitants in 27 European countries and Flanders (2006)

Although the number of road fatalities dropped significantly at the beginning of the 1990s, the trend has been less marked in recent years (as illustrated in Figure 2). In the battle for road safety, the European Union has set itself the ambitious goal to reduce the number of people killed in traffic between 2000 and 2010 by half (European Commission, 2001). In addition, challenging road safety targets have been set on national levels (see e.g., European Commission, 2006a; European Road Safety Observatory, 2006; Federale Commissie voor de Verkeersveiligheid, 2007; Organisation for Economic Co-operation and Development and European Conference of Ministers of Transport, 2006). However, if the trend continues at the same rate, according to the European Commission's Mid-Term Review (2006a) 32,500 people will die from road accidents in 2010. In order to achieve the EU-25 target of 25,000 fatalities at most by 2010, some additional effort will probably be needed. In order to decide on which measures to take, the main factors leading to accidents and casualties need to be studied.

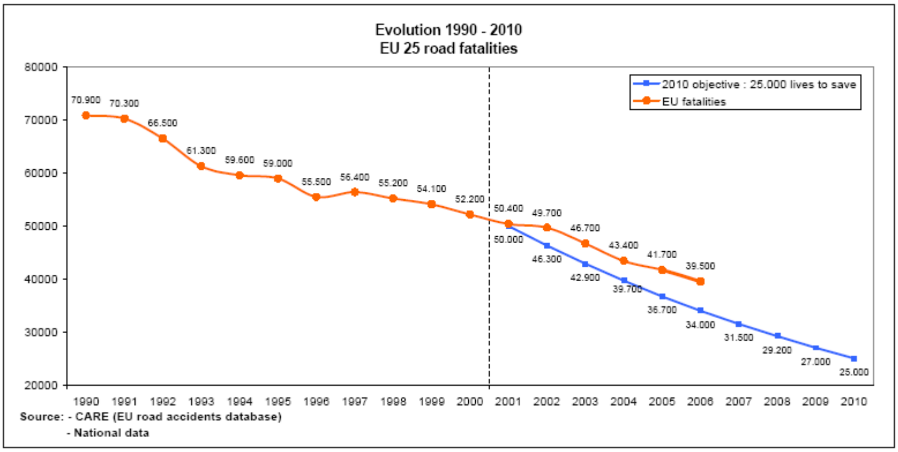


Figure 2: Evolution of road fatalities in the EU-25 (1990-2010) (European Commission, 2006a)

1.4.2 Decomposition

The former section indicated the size of the road safety problem on a worldwide, European and Belgian/Flemish scale. The occurrence of accidents and casualties results from a very complex mechanism. To enhance the level of road safety, it is essential to gain a clear insight into the underlying factors. In the past, many studies trying to explain why accidents occur have been conducted and various models have been developed aiming to describe the road safety situation in a country; to assess the impact of several influencing factors and measures; and to forecast the evolution in road safety (Christens, 2003; European Commission, 2004a; Hermans et al., 2006c; Organisation for Economic Co-operation and Development, 1997; Raeside and White, 2004; Van den Bossche and Wets, 2003). Many studies, for example Elvik and Vaa (2004), found risk factors leading to an increase in accident frequency or severity. The goal in safety analysis is to examine factors associated with accidents in order to identify the factors that can be changed by countermeasures to enhance future safety (Evans, 2004). Significant determinants of road safety will be described below.

The remainder of this section is divided into four parts. First, the human-vehicle-infrastructure decomposition is explained, followed by the exposure-risk decomposition. Thirdly, the C3-R3 systems approach is discussed and this section concludes with a description of the road safety risk domains for which performance indicators will be selected.

1.4.2.1 Human-vehicle-infrastructure

Road safety problems have traditionally been viewed as the result of malfunctions in the road transport system. This road transport system consists of three components: the road user, the vehicle and the infrastructure (European Commission, 2004a). Each accident is in most cases a direct consequence of failure in one or several of these three factors who influence each other: the desired behaviour of the road user (e.g., no drunk driving), the intrinsic safety of the vehicle (e.g., no technical defects) and the intrinsic safety of the infrastructure (e.g., good road surface conditions). The human factor is considered to be the most contributory one (Sabey and Taylor, 1980).

The road transport system comprising these three components is integrated in a broader environment, where it interacts with other factors. In numerous national and international studies the impact of various factors on road safety has been assessed across countries and/or over time (e.g., Eisenberg, 2004; Fridstrøm et al., 1995; Hakim et al., 1991; Hermans et al., 2006a; Melinder, 2007; Scuffham, 2003; Van den Bossche et al., 2005). From literature we deduce the following influencing factors of road safety closely linked to the human-vehicle-infrastructure system:

1. Regulation: laws related to alcohol and speed, safety actions and policies, ...
2. Demography: age distribution of the population, family composition, ...
3. Economy: unemployment rate, gross national product, income level, ...
4. Culture: religion, ...
5. Climate: precipitation, snow, sun, ...

1.4.2.2 Exposure-risk

Countries are often compared in terms of risk. In (SafetyNet, 2005a) risk is defined as the ratio of road safety outcome and amount of exposure (Eq. 1-1).

$$risk = \frac{road\ safety\ outcome}{exposure} \tag{Eq. 1–1}$$

The number of accidents and the number of casualties are two common examples of road safety outcomes. Exposure measures the degree of participation in traffic. There are several – traffic as well as persons at risk – exposure estimates (detailed information is given in SafetyNet, 2005a). The size of the population (number of inhabitants), the vehicle fleet (number of registered vehicles), the length of the road network (number of kilometers) and the driver population (number of driving licenses) are available and often used measures. However, they do not always act as good measures of exposure. This is due to differences in socio-economic conditions, population density, vehicles per citizen and transport mode split between countries (Al Haji, 2005). Other possible exposure measures are the number of trips or the time in traffic. In general, the number of vehicle (or passenger) kilometres travelled in an area (country or region) during a certain period of time (a year, a month, a day) is considered the most relevant measure for exposure. Unfortunately, this information – in the required level of detail and on a systematic basis – is usually lacking (SafetyNet, 2005a). Exposure data are often collected by counting traffic, travel surveys, local exposure measurements and indirect exposure estimates (e.g., based on fuel sales) (Organisation for Economic Co-operation and Development, 1997).

From (Eq. 1-1) it can be seen that road safety outcomes can be decomposed into risk and exposure. As shown in (Eq. 1–2), the expected number of accidents is equal to the product of exposure and risk. This type of risk is referred to as accident risk.

$$\#\ accidents = exposure \times accident\ risk \tag{Eq. 1–2}$$

In case we express road safety by the number of casualties, a third component appears in the formula (Eq. 1–3), namely the consequence which expresses the probability of getting injured once an accident occurred (Rumar, 1999). In this manuscript, the terms 'consequence' and 'injury risk' refer to the same.

$$\#\ casualties = exposure \times accident\ risk \times consequence \qquad \text{(Eq. 1–3)}$$

In principle, there are three ways for reducing the number of casualties: the amount of travel or the traffic volume can be reduced (e.g., promoting public transport); the accident risk can be reduced (e.g., driver training); and the level of severity in case an accident happened can be reduced (e.g., airbag). Such a decomposition enhances our knowledge since more information about the underlying road safety aspects becomes available. Each component – exposure, accident risk and consequence (or injury risk) – is influenced by a possibly different set of factors (Hermans et al., 2006b; World Health Organization, 2004). Economic and demographic factors affect exposure, while excessive speed and drugs are examples of factors that influence accident involvement. Helmet wearing and trauma management have a positive impact on (post-)accident injury severity.

1.4.2.3 C3-R3 systems approach

The previous sections decomposed road safety in human-vehicle-infrastructure on the one hand and exposure-risk on the other. In this section, a systems approach is presented. The systems approach focuses on the relationships and dependencies between the various individual elements of the transport system. It identifies the different stages at which policymakers can intervene to promote road safety. Zein and Navin (2003) state that effective road safety measures can merely be achieved by understanding the links between the elements of the system.

The first attempt to define road safety from a systems perspective occurred using the methods of Dr. Haddon (Zein and Navin, 2003). The three transport system components (road user, vehicle and road system) and three temporal sequences (before, during and after the crash) were combined into a Haddon matrix containing nine cells. Each cell identified the areas in which interventions could be initiated. For example, possible interventions aimed at the human factor are education (before the crash), restraints (during the crash) and trauma management (after the crash). We refer to Haddon (1980) for more information on this subject.

A further elaboration of the Haddon matrix resulted in 2003 in the presentation of the CR-R3 systems approach of Zein and Navin. The fundamental building blocks of the C3-R3 approach are:

1. three entities: the road user, the vehicle and the road system
2. three pre-crash timeline phases: creation, cultivation and conduct (which affect accident frequency)
3. three post-crash timeline phases: response, recovery and reflection (which affect accident severity)

Every combination of entity and timeline phase represents a cell in the C3-R3 system. This approach is visualised in Figure 3.

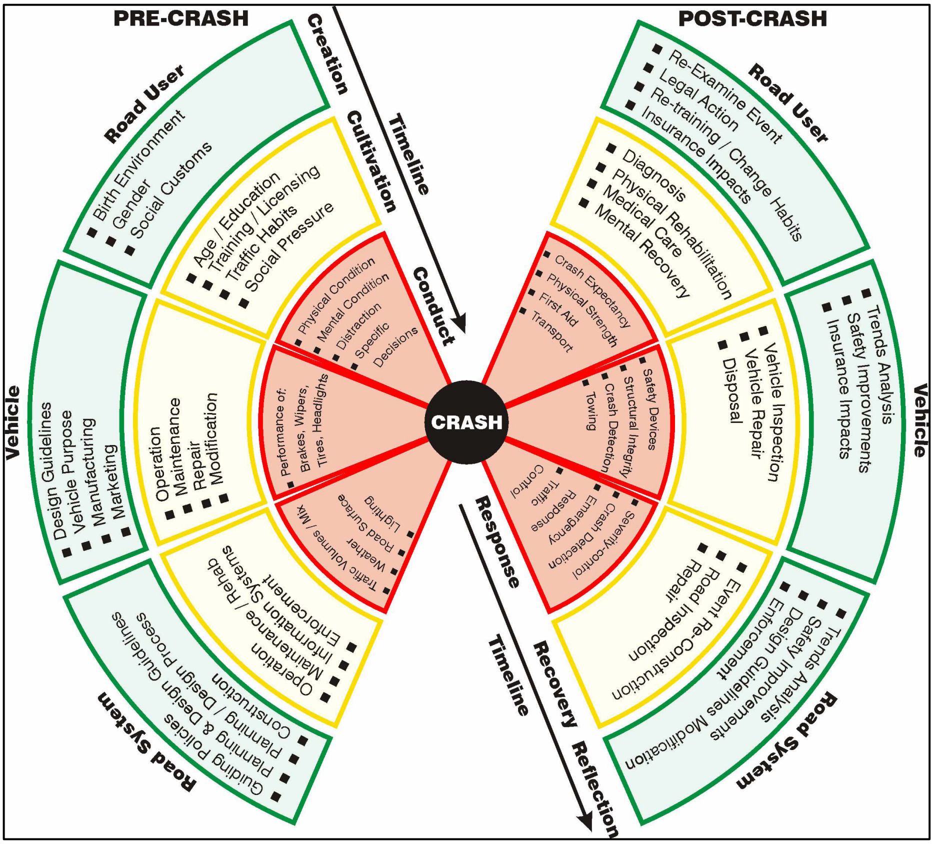


Figure 3: The C3-R3 systems approach (Zein and Navin, 2003)

1.4.3 Road safety risk domains

It can be inferred from the previous sections that road safety is a complex matter. Many factors have an influence in various ways. In this research, the overall road safety performance of countries will be assessed taking the most important risk domains into account. This importance can be described in terms of the next three aspects (European Transport Safety Council, 2001):

a) a stronger relationship with road safety

b) a larger contribution to accidents

c) to be influenced by measures[4]

Based on these three aspects as well as main road safety problems stated in national road safety plans (e.g., Begeleidingscomité, 2001) and relevant road safety indicators literature (European Transport Safety Council, 2001; Vis, 2005; World Health Organization, 2004) the following risk domains for road safety are proposed:

1. alcohol and drugs
2. speed
3. protective systems
4. daytime running lights (DRL)
5. vehicle
6. roads
7. trauma management

These road safety risk domains are discussed in detail in Chapter 3. At this point, we state that they can be considered to be central themes of road safety which can lead to a significant improvement in the level of road safety. Of course, other risk domains could be considered as well. However, based on a review of safety policies in the European Union and its member states, the European Transport Safety Council (2001) suggested the development of safety performance indicators related to behaviour (i.e., speed, alcohol and seat belts), vehicles, road and trauma management. On the basis of this report, the work package on safety performance indicators of the SafetyNet project (Vis, 2005) selected seven SPI domains to investigate. The research described in this

[4] This last aspect, policy impact, eliminates some factors such as the weather and the demography.

manuscript starts from the set of road safety risk domains used in the SafetyNet project (Vis, 2005) as it is the most recent source with respect to road safety performance indicators.

To conclude, the association of the seven risk domains with the 'human/vehicle/infrastructure' decomposition on the one hand and 'exposure/accident risk/injury risk' on the other is briefly discussed. First, alcohol and drugs, speed, protective systems and daytime running lights refer to the human behaviour. Other examples are distance-keeping, usage of crossing facilities and running a traffic light. The fifth domain is clearly linked to vehicle while the sixth domain represents infrastructure. In case the surroundings of an accident are looked at in a broader sense, the seventh domain – trauma management – is also linked to the third component of the decomposition.

Second, the relationship of the seven risk domains with accident risk and/or injury risk can be described (World Health Organization, 2004). Note that accident occurrence is linked to the pre-crash phases and accident severity to the post-crash phases of the C3-R3 approach. The domains alcohol and drugs, speed, vehicle and roads all both affect the probability of an accident to happen and the severity of the injury once the accident happened. Daytime running lights merely influences accident frequency whereas protective systems and trauma management have an impact on accident severity.

1.5 Road safety indicator frameworks

Like other policies, road safety policy could benefit from a collection of indicators able to measure changes and progress towards postulated objectives (European Environment Agency, 1999; Hens et al., 2005). In this section, indicator frameworks or conceptual models for structuring road safety indicators and their interrelationships are described. The concepts of 'road safety risk domains' and 'safety performance indicators' are presented as parts of a general road safety framework. In literature (see e.g., Environment Canada et al., 1999; Maclaren et al., 1995; Segnestam, 2002) several types of indicator frameworks can be found. Each framework organizes the information in a unique way. For example, in a goal-based framework indicators are organized according to how they correspond to various goals; a causal-based framework structures indicators into categories of pressure, state and response (or an extension hereof); and a domain-based framework ensures that the main components are taken into account (see e.g., Kaplan and Norton, 1992)[5]. With respect to road safety indicators, a causal-based framework (Section 1.5.1) and a goal-based framework (Section 1.5.2) will be discussed.

1.5.1 Causal road safety framework

The first framework (see Farchi et al., 2006; Lammar, 2006) presents the causal chain of road safety effects. Eight components are described for which indicators could be identified and evaluated.

- Driving forces: the principal driving forces are factors that create the need to travel. The degree of mobility is affected by the economic status of a country (e.g., employment rate), the distribution of wealth (e.g., average income), the distribution of population (demographic factors such as the number of inhabitants and the age distribution) and the physical geography of the country.
- Pressure: pressure factors result from the need to travel. The most important ones are cultural and social norms which create the interest in having a car and mainly travelling by personal transport.
- State: the state consists of several topics related to the conditions that influence the quantity of exposure. The degree of urbanization, the relative location of homes, schools, shops and work places, the age and quality of the vehicle fleet, the size and quality of the road network, the organisation of public transport and the climate are considered to be the most important factors.
- Exposure: exposure has already been discussed in Section 1.4.2.2. Two good ways to measure exposure is in terms of distance travelled and in terms of time spent in traffic. These quantities can be disaggregated by road user mode and road type. By means of activity based modelling (see e.g., Janssens and Wets, 2005), more detailed exposure information (e.g., travel route,

[5] The balanced scorecard approach can be seen as an example of a domain-based framework in which an organization is looked at from four perspectives (i.e., customer perspective, internal business processes, employee perspective and financial perspective).

travel mode choice and trip purpose) can become available, rendering the analysis more realistic. Exposure is affected by driving forces, pressure and state components (e.g., economic factors and urban population density) and influences the subsequent components.

- Accident risk: the concept of risk has been discussed earlier (see Sections 1.4.2.2 and 1.4.2.3). Factors affecting the probability of an accident are sometimes referred to as primary risk factors. Secondary risk factors influence the level of injury in case an accident happened. Some factors affect both accident frequency and severity. Examples of primary risk factors are drinking and driving, speeding, use of mobile phone, auditory or visual disturbance, active safety of vehicles, new driver, older road user, children without supervision, tiredness, medical disorder, infrastructural design and maintenance.
- Injury risk: examples of factors affecting the level of injury are drinking and driving, speeding, usage of seat belts, helmets and other protective systems, passive safety of vehicles, young and older road users, infrastructural aspects such as crash-protective roadsides and barriers, quality level of the rescue and pre-hospital emergency care and the health care system, etc.
- Effects: the former components result in health effects, namely accidents and casualties. From the registered accident information, the number of injury accidents, the number of fatalities and the number of injuries can be determined. Other possible effect factors are the years of life lost, the degree of invalidity and the psychological effect.
- Actions: actions include a wide range of preventive interventions, policies, laws, structural changes, etc. These actions mainly relate to engineering, education and enforcement. The effect of actions on the target can be monitored by means of indicators.

In Figure 4 it is shown that driving forces such as the economic, demographic and geographic situation exert pressure on our society leading to a certain state in terms of the size and quality of the vehicle park and the road network. This influences the amount of exposure, expressed in kilometres travelled or hours spent in traffic. The probability of getting involved in an accident (given a certain level of exposure) is affected by primary risk factors such as speed. Secondary risk factors for their part have an impact on the severity of injury once an accident happened. Several risk factors exist. In this research, we focus on the seven road safety risk domains listed in Section 1.4.3 and presented in Figure 4. The risk ends in road safety outcomes or effects like the number of fatalities. The last component of the causal framework consists of road safety enhancing policy measures. Actions can be targeted more or less directly (i.e., changing exposure or reducing accident and injury risk will result in fewer accidents and casualties) or indirectly (i.e., by trying to alter a particular driving force, pressure or state factor).

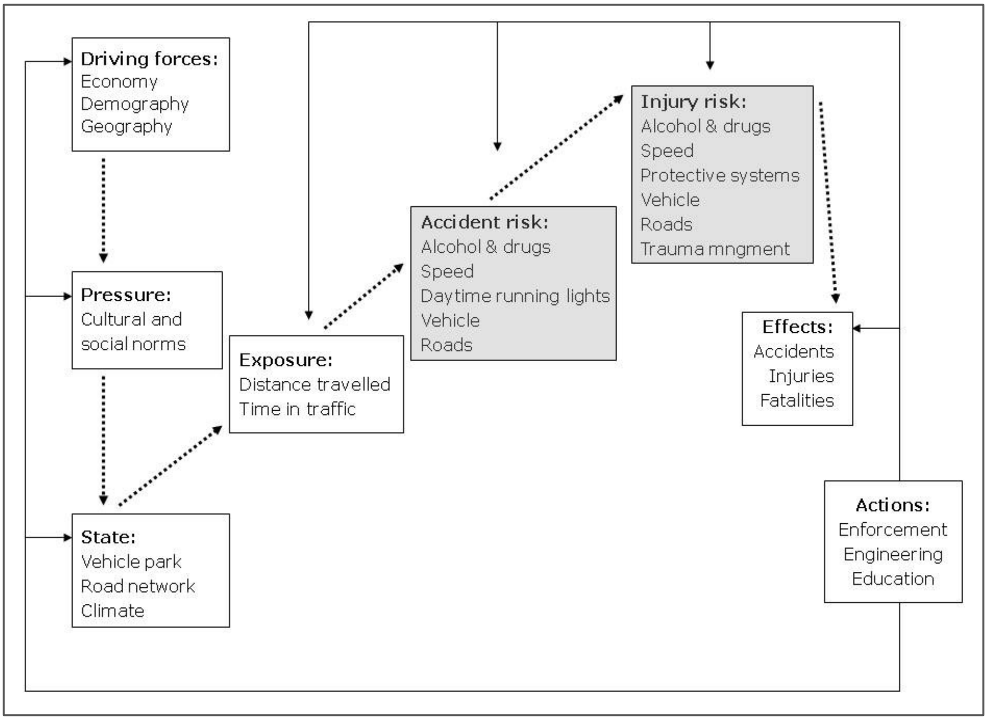


Figure 4: Causal road safety framework

1.5.2 Target hierarchy for road safety

In the second road safety framework indicators are specified with respect to various goals (e.g., reducing final outcomes). In 2000, the New Zealand's road safety target hierarchy originated (National Road Safety

Committee, 2000). This concept has been adapted in the European SUNflower project (Koornstra et al., 2002; Morsink et al., 2005) and was used in the European SafetyNet project as well (Vis, 2005). In fact, three dimensions should be considered (Morsink et al., 2005). The vertical dimension (shown in Figure 5) consists of five levels of the pyramid. At the horizontal level, road safety problems can be specified in a disaggregated way such as per road user group, transport mode, road type or region. The third dimension is time allowing to show the development of factors in both the horizontal and vertical dimension over time. Each component of the vertical dimension will be discussed below.

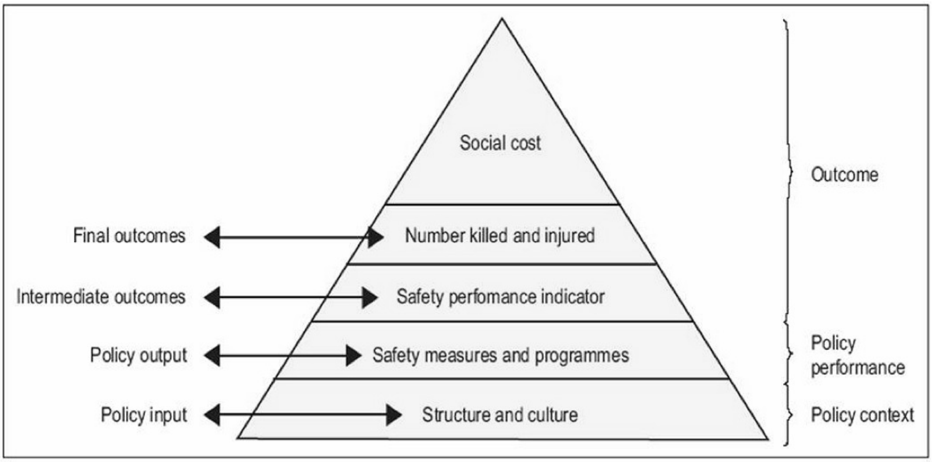


Figure 5: Target hierarchy for road safety (Morsink et al., 2005)

Social cost, at the top level of the pyramid, is the aggregated measure of all costs that accidents impose on the community (National Road Safety Committee, 2000). In order to determine the overall cost several cost components need to be taken into account. Economic costs do not reflect the pain, loss of function, emotional stress and other suffering to the casualties and immediate families (Evans, 2004). The Federal Highway Administration (Cambridge Systematics Inc. and Meyer, 2008) considers the following costs: property damage, lost earnings, lost household production, medical costs, emergency services, travel delay, professional rehabilitation, workplace costs, administrative costs, legal costs and pain and lost quality of life. In 2005 dollars, this resulted in a per person cost of \$3,246,192 for a fatality and \$68,170 for an injury. The European

Commission introduced the rule of thumb of €1,000,000 per fatality (Despontin et al., 1998). Later, €3,600,000 has been quantified, taking immaterial costs into account as well. For Belgium, the marginal unit value of preventing a road casualty is estimated at €2,004,799 per fatality, €725,512 per seriously injured and €20,943 per slightly injured person (De Brabander and Vereeck, 2007). Empirical data on human and economic production losses as well as on direct accident costs such as medical costs, hospital visiting costs, accelerated funeral costs, property damage, administrative costs of insurance companies, lawsuit costs, police and fire department costs and congestion costs were used in this computation. The overall monetary outcome depends on the final outcomes at the level below.

Final outcomes/number killed and injured consist of the number of casualties and need to be as low as possible. They can be further described in terms of road user age, transport mode, location and type of accident (Luukkanen, 2003). Road safety targets may be defined as an absolute or relative level of road safety. Targets in absolute terms are more widely understood. However, by presenting final outcomes as fatality or mortality rate instead of absolute numbers, changes in mobility respectively population are taken into account. Rates per person (e.g., mortality rate), per vehicle (e.g., fatality rate) or per unit of travel (e.g., fatality risk) are often used.

Intermediate outcomes/safety performance indicator are situated at the middle level. Certain road safety risk domains are responsible for the occurrence of accidents and casualties. The indicators at this level are called safety performance indicators. Thereby, these indicators provide the link between the final outcomes and the policy output. Intermediate outcomes are measured because they are generally reliable indicators of how well road safety interventions are working (National Road Safety Committee, 2000).

Policy output/safety measures and programmes refer to the nature of national road safety plans, action programmes and safety related standards and legislation. Examples are the number of police patrols, the budget spent on road safety campaigns, the legal speed limit on various road types, the probability of getting caught for drinking and driving and the penalty level of seat belt violation (Morsink et al., 2005; National Road Safety Committee, 2000).

Finally, **policy input/structure and culture** refer to the policy context, such as public attitudes towards risk and safety, the organisation of a country and its historical and cultural background (Morsink et al., 2005).

In the target hierarchy, not one but many targets regarding road safety can be set (National Road Safety Committee, 2000). Social cost and final outcome targets are the headline targets designed to capture attention. Nowadays, final outcome targets are widely used in national, regional and local road safety strategies and programmes (European Road Safety Observatory, 2006). Intermediate outcome targets (e.g., a particular seat belt wearing rate or speed violation share) are of interest to road safety professionals as they reveal the effect of individual interventions. Policy output targets (such as the number of police controls in relation to speed) can point out how well the postulated work programme is respected whereas policy input targets can be formulated in terms of attitude (e.g., the proportion of drivers in favour of extra speed cameras).

The road safety target hierarchy presented here is an interesting and valuable framework. On all levels, indicators could be defined, targets set and values compared. In fact, even countries showing similar final outcomes may differ to a considerable extent (Koornstra et al., 2002; Morsink et al., 2005). Their specific risk domains, policy, background, norms, etc can be captured in different indicator values.

1.6 Road safety performance indicators

So far, this chapter gave an introduction to indicators and the road safety problem. Moreover, two road safety indicator frameworks were presented. Keeping the major objective of this research in mind – i.e., developing a scientifically sound and appropriate methodology for the creation of a composite road safety performance index that can be used for the monitoring and comparison of road safety performance among countries – the remainder of this manuscript focuses on road safety performance indicators. More specifically, relevant indicators for the accident risk and injury risk factors presented in Figure 4 and the intermediate outcome level in Figure 5 will be identified and combined in a road safety performance index. Nevertheless, the index methodology presented in this manuscript might be of value to combine other types of road safety indicators as well.

The choice of performing research on road safety performance indicators has been influenced by a number of aspects. Traditionally, road safety research has been based on accident data. However, simply counting accidents and casualties gives an incomplete indication of the level of road safety (Al Haji, 2003). Safety performance indicators help in understanding the processes that lead to accidents. The purpose of SPIs is threefold: to reflect the current safety conditions of a road traffic system; to measure the influence of safety interventions; and to compare different road traffic systems such as countries (Vis, 2005). SPIs can give a more complete picture of the level of safety, give direction to policy instruments and can point to the emergence of new problems at an early stage before these problems show up in the form of accidents (European Transport Safety Council, 2001; Luukkanen, 2003). Because of the high information density they allow quicker and more detailed analyses and monitoring than accidents do (European Transport Safety Council, 2001).

As is the case for all indicators (see Section 1.2) road safety performance indicators can be used as a tool in policy analysis and communication. The current state and trends in road safety performance can be described and compared (in space and over time). Moreover, these indicators can indicate the success of countermeasure programs and support policy decisions regarding existing and new measures (National Research Council, 2002). Finally, they can be used to formulate specific road safety performance targets and identify priorities.

1.7 Structure of the manuscript

The previous sections already pointed out the value of determining road safety performance indicators. The relative road safety performance of countries will be assessed based on a composite road safety performance index. Indicators representing main road safety risk domains are selected and combined thereby presenting an enriched picture of road safety. The advantages (and disadvantages) of combining indicators in an index and the methodological process underlying the creation of a composite road safety performance index are described in the next chapter (Chapter 2).

To obtain an index a number of steps which are in fact small research domains in itself need to be gone through: selecting relevant indicators to combine, collecting indicator data, performing data analysis, weighting and aggregating the individual indicators, testing the robustness of the road safety index and computing, evaluating and visualising the final results. Each step is the topic of a particular chapter. In general, this manuscript deals with aspects such as: how can good road safety performance indicators be selected; are the indicators appropriate for combination; which techniques are available to equalize the different measurement units of the indicators; how to assign weights to each indicator; is total compensation between good and bad indicator scores acceptable; how robust is the constructed index; and how to interpret and visualise the results? From literature, relevant options will be listed and evaluated. In the end, a final index methodology will be presented resulting in a road safety performance index score for each country. Based on these scores, the relative rank of a country in terms of road safety performance can be determined. Contrary to the traditional road safety research, this rank takes several road safety performance aspects into account rather than one final outcome indicator.

Chapter 3 deals with the first step in the creation of a composite road safety performance index, i.e., the selection of road safety performance indicators. Each of the seven road safety risk domains – alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management – needs to be quantified by a set of indicators. Therefore, based on literature, existing indicators will be listed. Next, the possible indicators are evaluated based on several selection criteria. This chapter results in a set of best needed and best available road safety performance indicators.

To be able to illustrate and discuss the methodology for creating a composite road safety performance index, indicator data will be collected next (Chapter 4). For each risk domain, relevant data sources will be identified and data for the best available indicator collected. To be able to construct a road safety performance index all indicator values should be available for the same set of countries and relate to the same period of time. Moreover, before any analysis can be done the direction of the indicators should be assessed and geared to one another (e.g., a higher indicator value is expected to imply fewer casualties). Since the concept of road safety performance indicator is relatively new, some data limitations might need to be dealt with. At this moment, data will be collected with respect to seven risk domains for a large set of European countries relating to a period as recent as possible.

On the collected indicator data set some first analyses are performed in order to get more insight into each indicator separately (univariate analysis) as well as into the structure and interrelationships of the whole indicator set (multivariate analysis). Based on the analyses in Chapter 5 it should be concluded whether the indicator set is appropriate for combination in an index. The univariate analysis consists of three aspects. First, the data will be visually presented and described by means of summary statistics. Second, data transformation (normalisation) techniques rendering the data comparable over the indicators are discussed and third, the issue of missing values is briefly handled. Subsequently, multivariate analyses are performed in order to get an idea about the degree of association between the indicators, about the internal consistency of the indicators, about which indicators can be grouped together, about (dis)similar performance between countries and about the explanatory power of the selected indicators in terms of road safety output.

In Chapter 6, the issue of assigning a weight to each indicator is handled. From literature, several weighting methods that are commonly used and possibly interesting for the road safety case will be deduced. In addition to assigning an equal weight to each indicator, expert opinions can be taken into account or weights could be deduced from the data. A description of each weighting method and its advantages, disadvantages and requirements are given together with the results from applying each method. The methods will be evaluated based on these characteristics.

The topic of Chapter 7 is an optimization model which can be used for finding the most optimal road safety score for each country considering both intermediate and final outcome indicators. More specifically, based on an input-output data envelopment analysis model taking relevant road safety characteristics of each country into account, benchmarks can be obtained for each underperforming country. Using this information, country-specific road safety enhancing recommendations can be made (e.g., the risk domains requiring urgent action).

Chapter 8 discusses the process of aggregating indicators into an index. The mathematical formula or operator used for combining indicators needs careful thought. Once more, different options should be weighed against each other. Rather than only considering linear aggregation (or the arithmetic mean operator) and geometric aggregation (or the geometric mean operator) this chapter will start from the aggregation research field and its four broad classes of aggregation operators to deduce interesting ways of aggregation (such as by means of ordered weighted averaging operators). The degree of compensation allowed between good and bad scores as well as the incorporation of the aggregation idea of experts or decision makers will be considered in this respect.

In Chapter 9, the robustness of the road safety performance index is studied. Given the multistage index process, the index developer is left with a number of methodological choices that might influence the end result (i.e., the countries' ranking based on the index scores). Performing uncertainty and sensitivity analyses is an essential part of the index process as it shows how robust the index is and which extra information would imply more robustness. The impact of the methods selected and the assumptions made during the index construction process, e.g., with respect to indicator selection, weighting method and way of aggregating, can be quantified.

Based on the previously acquired information a final index methodology will be presented and index scores for a set of European countries computed (Chapter 10). Furthermore, the relationship between the composite road safety performance index and related indexes and indicators will be investigated. Finally, the index process is closed with a visualisation of the main results.

Finally, Chapter 11 summarises the main conclusions with respect to the research presented in this manuscript and some directions for future research.

Chapter 2 The road safety performance index

2.1 Introduction

The previous chapter gave an introduction to indicators, road safety and road safety indicators. In the remainder of this manuscript we focus on road safety performance indicators. These performance indicators offer valuable information as they can be used as tools to guide future action as well as to warn for high numbers of road casualties. By comparing safety performance indicator values between countries the main risk domains (e.g., speed) can be identified. That way, specific measures (e.g., speed enforcement) can be taken before the risk would manifest itself in a high number of road fatalities.

In this chapter, the process of combining a set of safety performance indicators in one composite road safety performance index[6] is discussed. That way, an overall road safety performance score can be obtained for each country. In general, we aim to create a road safety performance index that can be used for ranking countries based on their index score and therefore being an effective communication tool; for benchmarking as the relative road safety performance of each country can be assessed and the best-in-class country can be revealed; for monitoring the evolution over time and making predictions; and for policy supporting purposes since targets can be set and measures justified. In this respect, a scientifically sound and appropriate methodology for the creation of the index is prerequisite.

Very limited research has been performed regarding the combination of road safety indicators. However, the creation of indexes in other domains offers valuable insights. During the last decade, indexes have been developed in various domains. Al Haji (2005) makes a division according to the type of developer. A large number of indexes are being created and presented with the cooperation of international organisations, for example the human development index – United Nations; the environmental sustainability index – World Economic Forum; composite leading indicators – Organisation for Economic Co-operation

[6] The term 'road safety index' refers to the same in this manuscript.

and Development; and overall health system attainment – World Health Organization. Some other international indexes were developed within or in cooperation with universities and research institutes, such as the growth competitiveness index – Harvard University; the ecological footprint – University of British Columbia; and the general indicator of science and technology – National Institute of Science and Technology Policy of Japan. A review of these and other indexes is reported in Saisana and Tarantola (2002), Sharpe (2004), Wackernagel and Rees (1997) and McArthur and Sachs (2001).

This introductory section closes with an overview of the topics that will be handled in the remainder of this chapter. In the following section (2.2), some advantages and disadvantages of creating an index are discussed. Next, Section 2.3 presents the methodology to construct a road safety index consisting of several individual indicators. First, the set of performance indicators to combine has to be decided on (Section 2.4). The collection of indicator data as well as some univariate and multivariate analyses are topics of Section 2.5. In Section 2.6, we elaborate on common methods for weighting and aggregating indicators. Subsequently, we discuss the robustness of the index which can be assessed by means of an uncertainty and sensitivity analysis (Section 2.7). The computation, evaluation and visualisation of the final index values is the topic of Section 2.8. Concluding remarks are given in Section 2.9.

2.2 Indexes versus indicators

In the indicator literature, a distinction can be made between aggregated indicators on the one hand (i.e., composite indexes) and non-aggregated indicators on the other (i.e., individual indicators). Two basic approaches are available for aggregating indicators (Sharpe, 2004). The first one is the monetary approach in which variables are expressed in monetary terms first and then simply added (the gross domestic product is an example hereof). The second approach is the composite indicator approach in which domains represented by a set of indicators are combined using weights.

In Saisana and Tarantola (2002) and Tarantola et al. (2004) the main pros and cons of composite indexes are summarised. An index is characterized by the following advantages:

- It can summarise complex or multidimensional issues in view of supporting decision makers
- An index is easier to interpret than trying to find a trend in many separate indicators
- It facilitates the task of ranking countries on complex issues in a benchmarking exercise
- It can assess progress of countries on complex issues over time
- An index reduces the size of a set of indicators or includes more information within the existing size limit
- It places issues of country performance and progress at the centre of the policy arena
- It facilitates communication with the general public (i.e., citizens, media, etc) and promotes accountability

At the same time, the following disadvantages should be considered:

- An index may send misleading policy messages if it is poorly constructed or misinterpreted
- It may invite simplistic policy conclusions
- It may be misused, e.g., to support a desired policy if the construction process is not transparent and lacks sound statistical or conceptual principles
- The selection of indicators and weights could be the target of political challenge
- An index may disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action
- An index may lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored

In case the methodological index process is sound and transparent, the creation of an index over a set of indicators is worthwhile. Of course, the set of individual indicators provides an enormous amount of information. Nevertheless, as different risk domains jointly affect the frequency and severity of accidents, it is valuable to study the set of performance indicators simultaneously and combine all indicator information in one index.

The creation of indexes has largely progressed in recent years. Indexes developed in other domains and thereby using a particular methodology have been studied as they are helpful in developing a sound road safety index. Some examples of composite indicators are the internal market index 2004 (Tarantola et al., 2004); the European innovation scoreboard 2004 (European Commission, 2004b); the European e-business readiness index (Pennoni et al., 2005); the technology achievement index (United Nations, 2001); the environmental sustainability index (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2005); the index of economic and social well-being (Salzman, 2003); and the meta-index of sustainable development (Cherchye and Kuosmanen, 2004).

The creation of a composite road safety performance index based on performance indicators is new. However, we will briefly elaborate on the road safety development index described in Al Haji (2003; 2005) as it is one of the rare studies related to this subject. In this study, nine dimensions are selected for explaining and predicting the accident situation in several countries: traffic risk, personal risk, health index, education index, vehicle safety index, roads situation index, road user behaviour, standard of living and urbanization. However, in our research, we only consider road safety performance indicators situated on the level of intermediate outcomes in the target hierarchy for road safety (see Figure 5). Moreover, Al Haji compared the road safety situation and trends between ten Southeast Asian countries and Sweden and used a rather simplified methodology for combining the different dimensions. Attention was paid to the normalisation and weighting of the indicators. Index scores were computed based on four approaches: simple equal average, principal components analysis, assessments from expert's opinions and assessments from literature and theory review. The methodology that we suggest for the road safety index is discussed in the next section.

2.3 Index methodology

The calculation of an index requires an extensive use of statistical modelling and analysis techniques. Since indexes have been and are still being developed in various research domains, several methodologies for combining indicators exist. In general, more or less the same issues are encountered: indicator selection, normalisation, missing data treatment, weighting, aggregation and robustness testing. Other aspects can be added and the sequence of the different steps may change between studies. Based on a number of researches (Nardo et al., 2004; Nardo et al., 2005b; Salzman, 2003; Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006) we suggest the following steps for the creation of a composite road safety index.

1. Selecting indicators: this first step decides on the indicators to combine in an index. First, possible indicators are listed, followed by an evaluation based on a set of selection criteria. This results in a set of best road safety performance indicators.

2. Collecting indicator data: data need to be found for the indicators deduced in the previous step. Ideally, time series data for the set of best needed indicators are easily accessible in reliable databases for numerous countries. However, in practice, best available indicators characterized by available data of acceptable quality need to be used. As the different road safety aspects require their own databases (i.e., one overall data source having information on all aspects is nonexistent), indicator data from several sources need to correspond to the same set of countries and the same time period. In addition, all indicator values should imply the same (e.g., a higher value being favourable in terms of road safety output).

3. Univariate analysis: based on the collected indicator values, each indicator is studied separately. By means of visualisation and basic summary statistics, an idea about the distribution of the indicator values can be obtained. Extreme values might need a closer look as they could become unintended benchmarks. Next, all data should be made comparable. This process is called normalisation. As the indicators might differ in magnitude, be expressed in different units or expose great variation with respect to the mean, they may cause bias in the index. Often used methods for normalisation are standardisation, rescaling and rank

numbers. Thirdly, the imputation of missing values in the data set is dealt with at this stage. Several possible methods can be considered in this respect, like mean substitution, regression imputation, expectation-maximisation imputation or multiple imputation.

4. Multivariate analysis: this step provides insight into the structure and interrelationships of the data set by studying the indicators simultaneously. The appropriateness of the indicator set for combination in an index is assessed by means of correlation analysis, internal consistency analysis, principal components analysis, cluster analysis and regression analysis.

5. Weighting: a weight should be assigned to each indicator. The set of weights has a large impact on the index scores. In literature, several weighting methods can be found, none of them being a priori the best technique. Weights based on statistical methods (such as factor analysis), participatory methods (e.g., budget allocation), optimization methods (like data envelopment analysis) and equal weighting are the most common techniques. Relevant methods for the problem under study should be evaluated.

6. Aggregation: the mathematical formula for combining the indicators needs to be selected. In this respect, it is important to decide how the index consists of its indicators and to which extent compensation between good and bad indicator scores is allowed. Arithmetic averaging can be tested as well as other aggregation operators.

7. Robustness testing: it is important to rigorously test the robustness of the index or the stability in the output (i.e., the countries' ranking) given small changes in the input (concerning assumptions and methodological choices). The impact of the indicators included, the imputed missing values, the normalisation technique chosen, the selected weighting method and the applied aggregation operator can be quantified by means of uncertainty analysis. Moreover, a sensitivity analysis can reveal which of these factors is the most influential one.

8. Computing, evaluating and visualising final index scores: taking all previously acquired information into account a final index methodology is decided upon. Using the resulting (imputed) normalised indicator data, weights and aggregation operator, a final index score can be calculated for each country. The relationship between the composite road safety performance index and other related indicators or indexes will be assessed. Finally, the results should be visualised in a clear way. In addition to a ranking, various types of graphs can be produced.

All methodological steps will be briefly described in the next sections (and in more detail in Chapters 3 to 10). 'Section 2.4 Indicator selection' gives information on the first step. 'Section 2.5 Data preparation' relates to steps 2, 3 and 4; 'Section 2.6 Weighting and aggregating' handles steps 5 and 6; 'Section 2.7 Robustness testing' corresponds to step 7 and the final step is discussed in 'Section 2.8 Computation, evaluation and visualisation'.

2.4 Indicator selection

The road safety performance index will combine a set of performance indicators. Chapter 1 already provided insight into the multidimensional road safety problem and the main road safety risk domains identified at the European level. Each domain will be represented by performance indicators. Possible indicators are listed and subsequently evaluated based on selection criteria in order to obtain appropriate road safety performance indicators. More specifically, in Chapter 3 a set of best needed as well as best available indicators will be deduced.

2.5 Data preparation

Several steps in the index methodology relate to data. In particular, for the selected indicators, data need to be collected (step 2), analysed in a univariate way (step 3) and by means of multivariate techniques (step 4). The three steps will be successively discussed.

2.5.1 Collecting data

Attention should be paid to the collection of reliable data because the validity, interpretability and explanatory power of the constructed index depends on the quality and completeness of the data. It is advisable to choose only data sets with sufficient coverage, data freshness and methodological consistency (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006). In Chapter 4, the collection of road safety performance indicator data is discussed.

In this research, we focus on Europe as we want to evaluate the road safety performance of countries with a comparable level of mobility development (i.e., countries characterized by a similar transport system and motorisation rate). Nonetheless, a broader analysis on a worldwide scale would be interesting but data issues will play an even larger role. Beside a spatial boundary, our data set is limited in time. As no data are available for all indicators over some period of time, the road safety index will be computed for one year only. The most recent year for which all indicator data are available will be focused on. Progress towards postulated targets and relative success rates are to be studied in future research.

Finally, before doing any analysis on the data set, it has to be guaranteed that all indicators point in the same direction as this is crucial in the creation of an index. In other words, a high indicator value should always imply more (or less) casualties.

2.5.2 Univariate analysis

2.5.2.1 Describing and visualising

On the indicator data set some descriptive analyses will be performed. For each indicator summary statistics can be presented, such as the mean, median, minimum, maximum, standard deviation and skewness (see Chapter 5). They will, in addition to a visual presentation of the data, provide insight into the distribution of the values over the countries. Moreover, extreme values might become clear. One has to consider whether they represent real values or suspicious individual values that were erroneously entered into the data set. In the latter case, the data are subject to misunderstanding and may seriously distort final conclusions (Bird et al., 2005). Several statistical procedures aimed at detecting and dealing with outlier and aberrant observations exist. For example, the most extreme values may be equalized to the 5^{th} or 95^{th}

percentile (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006).

2.5.2.2 Normalising

In theory, each indicator could be expressed in a different measurement unit and on a different scale. It is important to cancel out the differences in magnitude to ensure that no indicator dominates the final index score due to large raw values. Another objective of normalisation is to convey information about a country's road safety performance in an easy-to-understand and meaningful way (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006).

The normalisation method has to be chosen. From literature, several techniques can be selected, like ranking, standardisation, rescaling, distance to a reference, categorical scale, indicators above or below the mean, methods for cyclical indicators, percentage of annual differences over consecutive years, etc (Nardo et al., 2005b; Organisation for Economic Co-operation and Development, 2003). Each normalisation method may result in other index scores. The impact of the selected normalisation technique on the end result can be quantified in step 7. In general, the normalisation method should take the data properties as well as the objectives of the composite index into account (Pennoni et al., 2005).

2.5.2.3 Missing values

After describing, visualising and normalising, an unequal number of observations per indicator is handled next. We might come across some missing values in the data set. In case only a small share of the indicator information for a particular country is missing, imputed values could be used rather than deleting that country, thereby enlarging the data set. Nevertheless, missing data cause extra uncertainty in the index.

Several statistical methods exist for imputing missing values, each making some assumptions about the missing patterns (Little and Rubin, 2002; Schafer and Graham, 2002; Schafer, 1997). Contrary to case deletion, single imputation and multiple imputation techniques consider the missing data as part of the analysis and try to come up with a realistic approximation to obtain a complete data set. In contrast to single imputation, several (more than one) values for each missing value are provided in case of multiple imputation, thereby better representing the uncertainty due to incomplete information. Possible imputation methods are mean/median/mode substitution, regression imputation, expectation-maximisation imputation and Markov Chain Monte Carlo multiple imputation. Applying these methods requires knowledge and careful consideration of the strengths and weaknesses of the various techniques in light of the available data (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006). However, the imputation of missing values is outside the scope of this research and will not be illustrated at this moment.

2.5.3 Multivariate analysis

The final data related step involves performing multivariate analyses on the whole indicator data set. The following techniques can be applied to gain insight into the structure and interrelationships of the data set: correlation analysis, internal consistency analysis, principal components analysis, cluster analysis and regression analysis. Each topic is discussed below and in more detail in Chapter 5.

2.5.3.1 Correlation analysis

Correlation coefficients for each pair of indicators are computed. Based on that information two indicators having a high correlation coefficient can be identified meaning they are partly showing the same aspect of the phenomenon under study (Nardo et al., 2004).

2.5.3.2 Internal consistency analysis

Cronbach's alpha, the most common estimate of internal consistency of items in a model or survey, can be computed (Nardo et al., 2005b). It measures the portion of total variability in the sample of indicators due to the correlation of the indicators. Alpha lies between zero and one with a high value implying that the indicators are measuring the same underlying construct (Malhotra and Birks, 1999).

2.5.3.3 Principal components analysis

Principal components analysis is a useful tool when investigating the relationships between indicators. The goal of this type of analysis is to reveal how different indicators change in relation to each other and how they are associated (Nardo et al., 2005b). The indicators are grouped in a few principal components explaining a large part of the variance in the data. The optimal number of principal components can be deduced from the analysis output. Moreover, the components are often rotated to enhance the interpretability. That way, more insight is gained into the grouping of indicators.

2.5.3.4 Cluster analysis

Beside grouping indicators by means of principal components analysis, groups of countries can be identified. Cluster analysis refers to a rich collection of statistical classification methods used to determine similarities (or dissimilarities) of subjects in large data sets (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006). A hierarchical clustering method can be applied first from which an optimal number of classes can be determined resulting in both a relatively small number of classes and a high level of similarity. This number of classes can subsequently be used in a non-hierarchical clustering method (e.g., *k*-means clustering) to obtain a classification of countries in homogeneous groups with a low within-variance and a high between-variance. Not only does this analysis provide information on the grouping of similarly performing countries, the cluster centres on each indicator can be plotted, thereby showing the main differences between several classes of countries. A cluster analysis can highlight the good and bad aspects of each class. Moreover, within each cluster, countries have a better basis for benchmarking their road safety performance as the group members are similar with respect to the data used to classify them (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006).

2.5.3.5 Regression analysis

Fourthly, a regression analysis provides interesting information. Each safety performance indicator can be seen as an explanatory variable and for the road safety case, the number of fatalities per million inhabitants is a possible dependent variable. In case high indicator values represent good performance, the coefficients of the explanatory variables are assumed to be negative. The model output (e.g., the goodness of fit value) can provide information on the explanatory power of the indicator set.

2.5.3.6 Concluding remarks on multivariate analysis

The main aim of performing multivariate analyses is to assess the appropriateness of the indicator data set and gaining insight into its structure. In addition, multivariate techniques such as correlation analysis and regression analysis can be useful in terms of indicator selection as well. For example, to select a limited number of speed indicators from a larger set of appropriate indicators, the indicator that turns out to be significant with respect to the dependent variable and that has a low correlation coefficient with the other speed indicators should be preferred. In the end, the set of indicators needs to have an acceptable level of correlation, internal consistency and clustering and explanatory power. Based on the results of the multivariate analyses, the selection of other indicators may be required. Selecting new indicators is justifiable as long as they appear to be appropriate indicators according to the indicator evaluation process.

2.6 Weighting and aggregating

The indicators are combined in an index by assigning a weight to each indicator and applying an aggregation formula. In literature, several weighting and aggregation options exist, none of them being the best technique to use in all circumstances. First, we determine the weighting methods most often used in the construction of an index and valuable for the road safety context. Section 2.6.2 discusses the options with respect to aggregation.

2.6.1 Weighting methods

By studying other indexes we obtain an idea about the most popular weighting methods. The e-business readiness index 2003 (Nardo et al., 2004) explored three weighting schemes, namely equal weighting, budget allocation (qualitative method) and factor analysis (quantitative method). Al Haji (2005) used simple equal average, a principal components analysis, an assessment technique from experts' opinions and an assessment technique from literature and theory review. Other examples are described in Nardo et al. (2005b), i.e., analytic hierarchy process, public opinion, data envelopment analysis, unobserved components models and conjoint analysis. It is common practice to experiment with various weighting techniques and compare the results. Nevertheless, the reasoning behind the choice of a particular weighting technique is often lacking in a study and simpler methods are more commonly applied than complex ones. In Booysen (2002) it is concluded that no weighting system is above criticism. Therefore, a thorough evaluation of the most relevant weighting methods for the topic under study is essential. In general, weights can be chosen to be equal, can be determined statistically, can involve the opinion of persons (e.g., experts) or can be optimized.

Below, five common weighting methods – equal weighting, weights based on factor analysis, budget allocation, analytic hierarchy process and data envelopment analysis – are briefly explained. More information is given in Chapter 6.

- <u>equal weighting</u>: assigning the same weight to each indicator is the most simple method. This is the main reason why it is often used in indexes. Salzman (2003) states that this method reduces the subjectivity of weights, has an interpretive meaning and is transparent. On the other hand, the importance of an indicator is not reflected by its weight and there is a risk of double counting (Nardo et al., 2005b).

- <u>weights based on factor analysis</u>: in this case, weights are determined based on the factor loadings of each indicator. Indicators with a strong capacity of explaining the variation in the data are expected to receive a relatively high weight (Pennoni et al., 2005).

- <u>budget allocation</u>: a group of experts is asked to allocate a budget of N points over a set of indicators, where a higher part of the budget should be given to more important indicators. The share of the budget assigned to an indicator equals its weight. To obtain good results, a team of qualified experts needs to be found. The weights resulting from a particular expert could be considered for the analysis as well as a set of weights based on the opinion of all experts.

- analytic hierarchy process: this method uses experts' opinions as well. Each indicator is compared to another indicator and the expert decides which of the two is more contributing to the overall goal and indicates the intensity of the strength (Saaty, 1980). In case two indicators are considered to have the same contribution to road safety, a score of one is given whereas an extreme difference in contribution results in a score of nine. All scores are then transformed in weights summing up to one.
- data envelopment analysis: in this case, the weights are endogenously determined from the data set as to obtain the most optimal index score for a country. This means that higher weights are attached to the indicators on which a country performs relatively well (Cherchye and Kuosmanen, 2004). Moreover, restrictions regarding the share of each indicator in the overall index score can be incorporated in the optimization problem to obtain more realistic weights. This method can result in best country-specific weights as well as one set of indicator weights that is the same for each country (like other weighting methods do).

For the road safety index, the advantages and disadvantages of the five previously described methods are worthwhile investigating in more detail. For a profound evaluation of weighting methods we refer to Chapter 6.

2.6.2 Aggregation methods

As is the case for weighting, aggregation is a potential area of methodological controversy in the field of index construction (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006). Various aggregation options exist of which the selection should fit the purpose of the study and the subject being measured. In the handbook on constructing composite indicators (Nardo et al., 2005b) three aggregation methods are mentioned. Additive or linear aggregation sums the weighted indicators. Secondly, the non-compensatory multi-criteria approach is described. This involves the creation of an outranking matrix. Thirdly, geometric aggregation implies that the index score equals the product of the indicators to the power of their weights. Linear aggregation is the best-known way of aggregating. However, the suitability of this type of aggregation has been questioned (see e.g., Munda and Nardo, 2005).

Apart from the three aggregation methods suggested in Nardo et al. (2005b) aggregation is a very extensive research domain in which numerous types of operators exist. They are all characterized by certain mathematical properties and aggregate in a different manner. In general, aggregation operators can be roughly divided into four broad classes: averaging functions, conjunctive functions, disjunctive functions and mixed functions (Beliakov et al., 2007). The class of averaging functions possesses interesting properties and consists of ordered weighted averaging (OWA) operators, an often used and comprehensible class of operators that is worthwhile testing. Good and bad performances can be emphasized differently and the degree of compensation allowed between the two can be reflected. Road safety professionals can express their aggregation opinion in natural language (e.g., in case a country scores badly on more than half of the indicators, its final road safety index score should be very small). This guideline is then translated mathematically and index scores respecting the statements can be computed (Yager, 1996). By changing the parameters OWA operators can generate a wide spectrum of scenarios. The issue of selecting a suitable aggregation operator is discussed in Chapter 8.

2.7 Robustness testing

So far, different steps in the methodological process for creating an index have been described. In a scientifically sound and appropriate index methodology, the robustness of the index is assessed by means of uncertainty and sensitivity analysis. More specifically, the stability in the countries' ranking due to changed decisions with respect to normalisation, weighting and aggregation is quantified. A comprehensive description of uncertainty analysis and sensitivity analysis, applied to the technology achievement index is given in Saisana et al. (2005). In general, uncertainty analysis estimates the uncertainty in the output taking into account the uncertainty affecting the input factors. At the same time, sensitivity analysis is defined as the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. In their handbook, Saltelli et al. (2008)

present a step-by-step plan for performing a sensitivity analysis on a model (see also Chapter 9).

Generally, the following questions can be answered by testing the robustness of the index (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2006):

- how do countries' ranks relate under alternative methodological scenarios?
- which methodological step accounts for the largest variability in the ranking?
- which countries have the most unstable ranks and why?
- what is the optimal set of methodological assumptions for a country?

As in other studies (e.g., Nardo et al., 2004) it would be interesting to observe how much the countries' ranking is affected by a change in the chosen set of indicators. In case a data set is characterized by some missing values, the impact of the imputed values on the end result could be quantified. Ideally, all potential sources of uncertainty should be tackled: the inclusion and exclusion of indicators, the imputation technique, alternative normalisation techniques, different weighting methods and various aggregation options. In the end, we aim to get insight into the robustness of the index with respect to the methodological decisions taken during the index process and into the most influencing factors resulting in uncertainty. In case the weighting method results in a high level of uncertainty, a decision regarding this method should be justified and taken.

2.8 Computation, evaluation and visualisation

We reached the final step of the index methodology now. In the previous steps, the indicators to combine were identified, data collected, analysed in a univariate and multivariate way and weighted and aggregated. Moreover, the impact of several methodological choices on the end result was quantified and the factor(s) causing most uncertainty pointed out. Taking all information acquired so far into account a final index methodology is decided upon. The resulting index scores are discussed and related to the results of other indexes or indicators. The link between the road safety performance index and a final road safety outcome indicator, the results from the SUNflowerNext study (Wegman et al., 2008) and the corruption perceptions index (Lambsdorff, 2004) could be assessed (see Chapter 10).

Moreover, composite indicators should be able to quickly and accurately communicate a picture to decision makers and other end users (Nardo et al., 2005b). The previous steps in the methodology require a lot of research. In case no clear presentation of the results is offered, all previous efforts will not attain full use. Several types of graphs can present the index values and its components.

There are interesting ways to display index scores, from simple tabular tools to more complicated multidimensional graphs and interactive software (Nardo et al., 2005b). For illustrative purposes, several ways of presentation used in other studies will be given below. While elaborating the road safety index, the most optimal type of visualisation needs to be determined. In this respect, the simplest mode of presentation that avoids being misleading and that is in accordance with the research objectives needs to be chosen (Bird et al., 2005). More specifically, presenting performance indicator or index data intended for public release should have intuitive appeal. Indications of variability such as confidence or uncertainty intervals may involve some statistical complexity in their determination but are usually adequately represented in a simple way. Furthermore, in case the results are for immediate action or discussion graphical methods are typically best; in case further analysis and comparison is involved a table may be preferred. When indicator or index data are plotted over time, the evolution can be shown. In addition, targets set for the future or the performance of benchmarks could be indicated on the graph as well as warning lines providing signals for urgent action.

One of the most common ways of presenting information is by means of bar charts. Figure 6 shows an overall bar chart presenting the commitment to development index and its 7 underlying dimensions The best, average and worst performing countries can be detected.

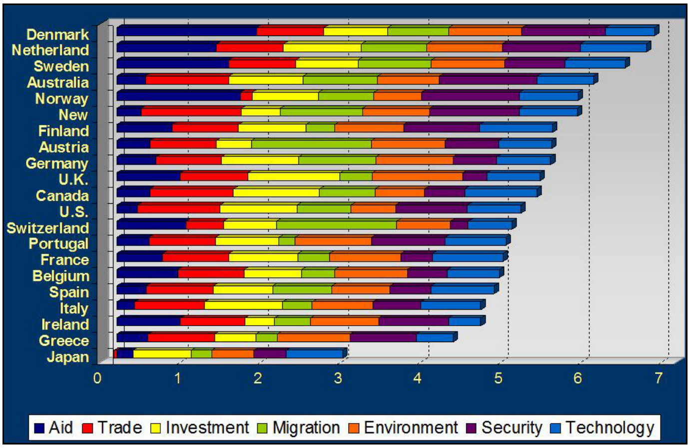


Figure 6: Overall bar chart for the commitment to development index (Roodman, 2005)

Spider diagrams or radar charts are a second group of interesting graphs. They can be used for presenting information on several aspects related to one item (e.g., percentage of road fatalities per transport mode or age class; protective systems wearing rate for seat belts, helmets and child's seats; different dimensions of the overall index; etc). Spider diagrams have multiple axes along which all the aspects can be plotted, as can be seen in Figure 7. Each variable is normalised between 0 and 100. In general, higher values, i.e., points near the edge, indicate good performance. To keep a clear overview, the countries in the data set are often divided into several subgroups in order to visualise the performance of a few countries with respect to each other or to the average of all countries.

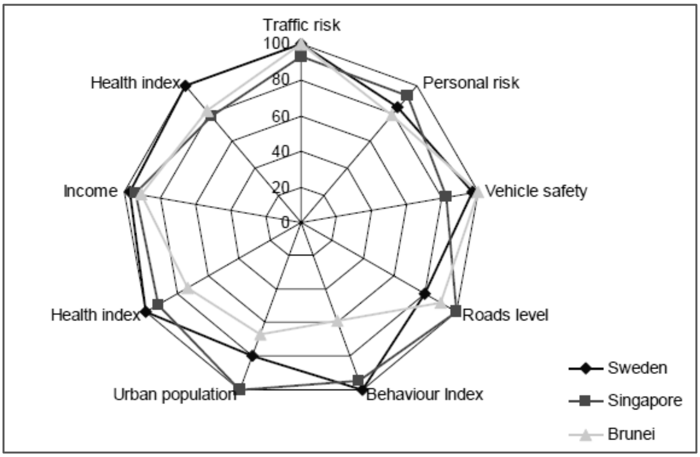


Figure 7: Spider diagram for 9 dimensions of the road safety development index (Al Haji, 2005)

Figure 8 shows the variability in rank for 23 countries with respect to technology achievement. Under different methodological assumptions, countries like Singapore or Norway obtain a totally different position in the ranking.

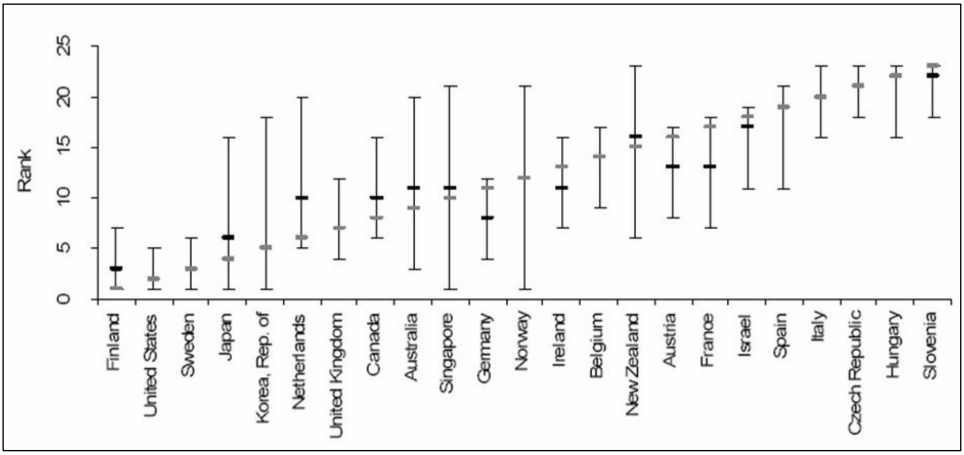


Figure 8: Ranking of countries based on the technology achievement index (Nardo et al., 2005b)

Next, presenting the performance of a country on two axes, the graph surface can be divided into four quadrants. In Figure 9, countries are positioned on their summary innovation index score (Y-axis) in relation to the average growth rate (X-axis). Countries in the upper right quadrant are doing best as they are moving ahead. On the contrary, countries in the lower left quadrant score badly on both dimensions, thereby falling further behind. The upper left quadrant countries are losing momentum while the lower right quadrant countries are catching up. Finally, a coloured map can be presented enabling the end user to see at a glance how a country is performing (see Figure 10).

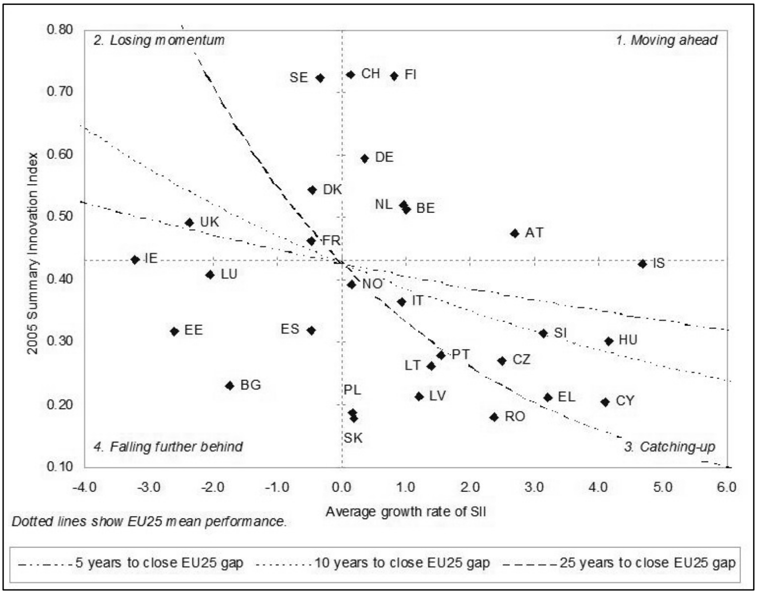


Figure 9: Four-quadrant graph for European innovation performance (Dupont, 2005)

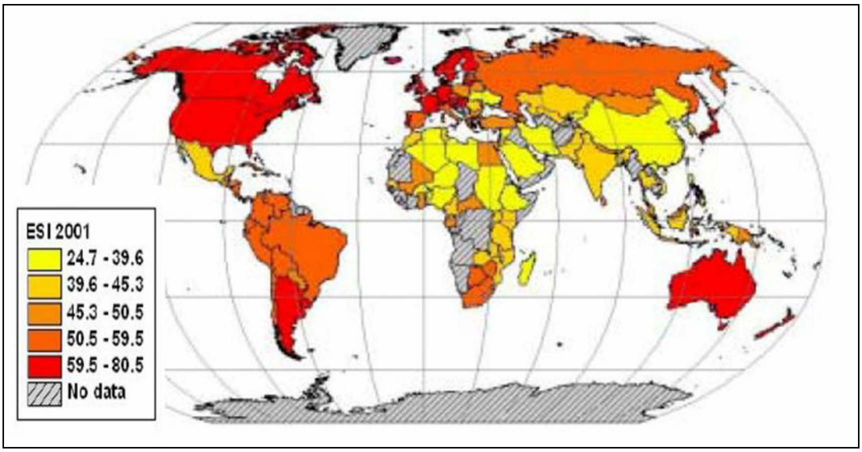


Figure 10: Coloured map for the environmental sustainability index (Yale, 2005)

The above graphs are useful tools for presenting road safety index results. Although road safety performance information is combined in one score, the decomposition in individual indicators might be of interest to some end users as well. The visualisation aims to indicate the relative performance of countries and the leaders and laggards. A (e.g., spider) diagram showing the performance of a country and its benchmark is useful to get insight into the aspects needing urgent action. A more detailed look into the policies and practices of the benchmark can provide useful inspiration (see e.g., SUPREME, 2007). Moreover, it would be worthwhile visualising changes in road safety performance over time.

2.9 Concluding remarks

The index process is a methodologically intensive one that needs considerable thought. Each step in this process will be thoroughly examined. More specifically, from literature (describing other indexes) relevant options for the road safety index case are selected and profoundly evaluated. The selection of a preferred method within each step will be supported by quantitative and qualitative assessment.

The results from applying each method will be compared against each other. The quantitative criterion used in this respect is the degree of correlation between the resulting countries' ranking (based on the index scores) and the final outcome ranking. Keeping the target hierarchy for road safety in mind (Figure 5) we want to create a composite index that has a clear link with the final outcome layer. That way, the composite road safety performance index and its indicators can act as warning signals before a specific risk factor would manifest itself in an increased number of accidents and casualties. In other words, the formulation of a performance index in accordance with the final outcome layer enables earlier and goal-oriented action. The reasoning behind the postulated link between the intermediate outcome and the final outcome layer is related to the use of leading indicators in the field of economics[7].

[7] Leading economic indicators help predicting the state of the economy in the future and provide policymakers with signals of turning points of economic activity (Gyomai, G. and Guidetti, E. (2008). *OECD Sytem of composite leading indicators*. OECD.

Although the quantitative criterion is something to hold on to in the appraisal of methods it is an insufficient basis for making sound decisions, given the dependency on the data used. Furthermore, to avoid the determination of the index methodology to become a purely mathematical exercise in which the highest correlation coefficient is the main point of interest, attention is paid to the qualitative evaluation of the different options. Consequently, the main advantages and disadvantages of each method will be discussed. Preference should be given to the method that offers additional insight to the end user. Primarily, the methodology should produce a valuable and acceptable index in terms of communication, benchmarking, monitoring and policy support.

In the subsequent chapters, each methodological step is handled in detail. Chapter 3 discusses the selection of indicators. In Chapter 4, indicator data are collected for a set of European countries. Univariate and multivariate analyses on the data set are the subject of Chapter 5. The issue of weighting and of aggregating is discussed in Chapter 6 respectively Chapter 8. The robustness of the index is assessed in Chapter 9 while Chapter 10 presents the final index methodology and computes, evaluates and visualises the road safety performance index scores. To conclude, the methodological process needs some revision from time to time as new information and methods may be worthwhile testing for the road safety case.

Chapter 3 Selecting indicators

The first step in the creation of an index is identifying the indicators to combine. In this manuscript, a composite index relating to the intermediate outcome layer (see Figure 5) is constructed. More specifically, the performances of each country with respect to seven main risk domains are combined in one overall road safety performance index score. More information on the risk domains is given in Section 3.1. These risk domains need to be represented by appropriate indicators. This implies that the 'best' indicators should be inferred from the set of possible indicators. In Section 3.2, selection criteria and the indicator card are described. The possible indicators and the evaluation results are discussed in Section 3.3. This chapter closes with concluding remarks on the selection of road safety performance indicators (Section 3.4).

3.1 Road safety risk domains

As described in Section 1.4.3 we start from the seven risk domains considered in the SafetyNet project (Vis, 2005) as it is the most recent source with respect to road safety performance indicators: alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management. Each risk domain is discussed successively.

3.1.1 Alcohol and drugs

Driving while being intoxicated causes a higher accident risk. A larger blood alcohol content (BAC) implies a higher probability of getting involved in an accident. More specifically, the relative risk of getting involved in an accident starts increasing significantly at a BAC level of 0.04 g/dl (World Health Organization, 2004). Moreover, in a number of studies (see Hakim et al., 1991) consumption of alcohol has been seen to increase the frequency of fatal accidents. Al Haji (2005) discusses a study (Thoresen et al., 1992) in which a positive correlation between the total number of fatalities in Victoria state (Australia) and alcohol sales was found and an inverse relationship with random BAC breath testing. Other findings (Hakim et al., 1991) are a negative relationship between both the number of accidents and the severity of injuries on the one hand and the minimum age for purchasing alcohol on the other. A study from the United States (Zador, 1991 in World Health Organization, 2004) shows that for single-vehicle accidents, each 0.02% increase in BAC level nearly doubles the risk of getting involved in a fatal accident.

According to the alcohol and drugs questionnaire of the SafetyNet team (no public source), accidents involving at least one driver impaired by alcohol caused 5.4% of the fatalities in Belgium in 2006. However, as an alcohol test is not always performed (from the questionnaire it appears that one third of the drivers involved in fatal accidents was tested), these figures are possibly an underestimation of the real size of the problem (Vanlaar, 2005). The Belgian Toxicology and Trauma Study (1997) estimated that 28% of all drivers who entered the emergency room after an injury accident had a blood alcohol content above the legal limit (Begeleidingscomité, 2001). During weekend nights, this percentage increased up to 50%.

In case of drugs, it is more difficult to quantify the impact as there is currently insufficient information about the concentration or combinations that may cause driving problems. In this respect, a distinction between medicinal and illicit drugs can be made. Moreover, the concentration of drugs is difficult to measure in a reliable way. However, it can be expected that drugs intoxication implies a higher risk.

3.1.2 Speed

Inappropriate or excessive speed has been identified as a highly important factor influencing both the number of accidents and the severity of injuries (see e.g., Elvik, 2005; Kweon and Kockelman, 2005). In high-income countries, excessive and inappropriate speed is a main cause in one third of the fatal and serious accidents (World Health Organization, 2004). The probability of becoming involved in an injury accident increases with a higher (average) speed and/or larger speed differences (Al Haji, 2005; Vis, 2005). The level of road safety could improve if the curve representing the speed variation is narrowed (i.e., a smaller difference in speeds and few drivers exceeding the speed limit) (Begeleidingscomité, 2001).

The probability that an accident will result in injury is proportional to the square of the speed; for serious injury proportional to the cube of the speed; and for fatal injury proportional to the fourth power of the speed

(Nilsson, 2004). In addition, the probability of a pedestrian dying as a result of a car accident augments exponentially as the speed of the car increases. Reducing vehicle speeds appears to have a significant effect on road casualties and pedestrian accidents (Balkin and Ord, 2001; Fridstrøm et al., 1995).

Factors that affect the choice of a particular speed are related to the road (width, layout, markings, quality of the surface), the vehicle (type, maximum speed, comfort), the traffic (density, composition, prevailing speed), the environment (weather, light, enforcement) and the driver (age, sex, reaction time, risk acceptance, alcohol level, car occupancy) (World Health Organization, 2004).

3.1.3 Protective systems

Protective systems play a role in case an accident has occurred as they determine the severity of the injury. Mandatory seat belt use has been one of the greatest success stories of road injury prevention and has saved many lives. Several empirical studies in the United States, Great Britain, Sweden and The Netherlands have shown that seat belt legislation, when followed by law enforcement, significantly reduces the number of fatalities and the severity of injuries (Hakim et al., 1991).

Not wearing a seat belt or helmet implies an increased injury or fatal risk. Several studies have found that seat belts, helmets and child's seats are very effective. The level of effectiveness depends on the impact speed, the type of collision and the position of the occupant. The use of seat belts reduces the probability of being killed in an accident by 40 to 50% for drivers and front seat passengers and by 25% for passengers in the rear seats as shown in Elvik and Vaa (2004). The world report on road traffic injury prevention (World Health Organization, 2004) shows the effectiveness of various protective systems. Wearing a seat belt reduces the risk of serious and fatal injury by 40 to 60%. In case motorised two-wheelers wear a helmet, fatal and serious head injuries are reduced by 20 to 45%. Furthermore, cyclist helmets diminish the risk on head and brain injuries (by 63 to 88%). Moreover, child's seats offer 70% reduction in the risk of dying for children under 1 year and 54% for children between 1 and 4 years (Begeleidingscomité, 2001).

3.1.4 Daytime running lights

Daytime running lights refers to motor vehicles having their headlights on during hours of daylight (SWOV, 2008). Daytime running lights help road users to better and earlier detect, recognise and identify vehicles (European Commission, 2006b). This increased conspicuity implies fewer accidents. The effect of daytime running lights on road accidents has been studied for several decades, starting in Northern European countries (see e.g., Elvik, 1993). In the handbook of road safety measures (Elvik and Vaa, 2004) a reduction in the number of multi-party accidents of 10 to 15% is linked to DRL use.

By 2006, in 14 member states there exists a regulation regarding daytime running lights i.e., Austria, Czech Republic, Denmark, Estonia, Finland, Hungary, Italy, Latvia, Lithuania, Poland, Portugal, Slovakia, Slovenia and Sweden. A report from Knight et al. (2006) concludes that the mandatory use of DRL in all member states would provide a net accident reduction. Recently, the European Commission has decided to introduce daytime running light on all new types of motor vehicles from the year 2011 onwards (Commission Directive 2008/89/EC of 24/09/2008).

3.1.5 Vehicle

Vehicles have become safer over time due to a progressively stricter regime of standards (European Environment Agency and Eurostat, 1999). Furthermore, there is a rapid evolution in vehicle technique and technology (Begeleidingscomité, 2001). Improvements in both active and passive safety resulted in a lower frequency and severity of accidents. Active safety features help the driver in avoiding an accident, such as anti-lock braking systems, traction control, driving aid systems and audible warning devices while passive safety features better protect occupants in the event of an accident, like frontal and side impact protection, airbags, load restraint and crush zones (Begeleidingscomité, 2001; National Road Safety Committee, 2000).

There is a link between vehicle age and risk. Occupants of a car produced before 1984 have approximately a three times higher injury risk compared to occupants of a newer car (World Health Organization, 2004). As the vehicle fleet is continuously being renewed to higher safety standards the presence of safety features in the overall vehicle fleet can be estimated by means of the age of the fleet.

Over time, crashworthiness, i.e., the protection that a vehicle gives in case of an accident, has improved. The new car assessment program evaluates vehicle crash performance by rating the vehicle models on their safety level for occupant protection, child protection and pedestrian protection. The European new car assessment programme EuroNCAP (www.euroncap.com) supplies information to consumers about the performance of new cars in crash tests since 1996. A higher EuroNCAP rank implies fewer fatal and severe injuries (European Transport Safety Council, 2001).

It is known that poor vehicle maintenance and technical conditions can also contribute to accidents (Al Haji, 2005). In high-income countries, the role of vehicle defects to accidents is around 3% (World Health Organization, 2004). Elvik and Vaa (2004) concluded based on a review of macro studies that there is no clear evidence that periodic vehicle inspection has an effect on the number of accidents. In Hakim et al. (1991), the authors refer to a study (White, 1986) concluding that the probability of accident involvement increases with the length of time between inspections.

3.1.6 Roads

The safety performance of the road transport system is the result of the combination of the functionality, homogeneity and predictability of the network, the road environment and the traffic involved (Vis, 2005). The road network influences accident risk as it determines how road users perceive the environment and offers instructions by means of signals (World Health Organization, 2004). Four influencing factors are safety awareness in the planning of new road networks, dealing with safety features in the design of new roads, safety ameliorations to existing roads and healing actions on locations with a high accident risk.

In addition, poor road surface conditions and defects in road design and maintenance contribute to an increased accident risk (Al Haji, 2005; European Transport Safety Council, 2001). Objects along the road provide a risk in case the road user gets involved in a (run-of-the-road) accident. The performance tracking of roads is the focus of EuroRAP (Lynam et al., 2004). The EuroRAP road protection score aims at assessing the degree to which roads protect against severe injury in case of an accident.

The road network consists of several road types. Despite the high speed allowed, motorways are considered to be the most safe type of roads. However, they represent only a few percentages (0 to 2.8%) of the total road network (European Union Road Federation, 2007). Rural roads account for a considerable share of all fatalities. The risk of being killed (per kilometre driven) is generally higher on rural roads than on urban roads and is four to six times higher than on motorways (Organisation for Economic Co-operation and Development, 2002). Rural road accidents are generally more severe than accidents on urban roads due to differences in operating speeds, road geometry, functionality, enforcement levels and other factors (Organisation for Economic Co-operation and Development, 2002).

3.1.7 Trauma management

Trauma management refers to the system responsible for the medical treatment of injured persons from a road accident. The probability of surviving and the quality of life after the accident are influenced by the level of trauma management (European Transport Safety Council, 2001).

Noland (2003) concludes that advances in medical treatment and technology have resulted in reductions in traffic related fatalities in all developed countries over time (1970-1996). Fatality figures are correlated with the level of medical facilities available in a country expressed in terms of population per physician and population per hospital bed (see Al Haji, 2005).

A report from the European Transport Safety Council (1999) concluded based on a review of European studies on traffic mortality that almost 50% of road fatalities die within a few minutes at the accident scene or on their way to the hospital. Concerning hospitalised casualties 15% dies within one and four hours after the incident while 35% dies after 4 hours. The following chain applies: actions on the accident scene, access to the emergency medical system, help by paramedics, medical care provided before arriving to the hospital, medical care in the hospital and rehabilitative psychosocial care (Hussain and Redmond, 1994 in World Health Organization, 2004).

3.1.8 Road safety risk measures

Seven risk domains have been discussed above. All domains will be expressed quantitatively by means of performance indicators. Ideally, each risk domain is represented by several indicators able to indicate the performance on a certain domain in a realistic way. We believe that the selection of appropriate indicators for the domains alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management provide policymakers with valuable road safety performance information. Specific measures can then be taken for tackling problem areas which in turn will result in fewer accidents and casualties and a reduced social cost (see Figure 5). Possible measures for each domain are listed below:

1. alcohol and drugs: alcohol limit, enforcement and measures against recidivism, campaigns
2. speed: speed limit, speed cameras, enforcement, sanctions, campaigns
3. protective systems: enforcement, sanctions, campaigns
4. DRL: regulation, campaigns
5. vehicle: design and maintenance standards, informing potential buyers, motor vehicle taxation
6. roads: safer road designs, hot spot analysis, road safety audits, road inspections
7. trauma management: investments in medical services and material, easy access to the accident scene

For each of the seven risk domains – having a strong relationship with road safety, contributing to accidents and which can be influenced by measures – a list with potential indicators will be formulated, one indicator being more appropriate than another. First, attention is paid to the selection process.

3.2 Selection process

In this chapter, the choice of a set of appropriate road safety performance indicators is discussed. This indicator set will contribute to the establishment of a monitoring system in order to (Farchi et al., 2006):

- measure the road accident phenomenon, its determinants and the trends
- facilitate the planning, monitoring and evaluation of programs and actions
- provide organisations and policymakers with valuable information to make comparisons and support their policies

Choosing appropriate indicators is not an easy task. For every risk domain, the strengths and weaknesses of each possible indicator need to be assessed. In this respect, the use of an indicator card (see Section 3.2.2) summarizing all relevant information is beneficial. That way, the various indicators can be judged based on the same aspects. First, attention is paid to one particular part of the indicator card, i.e., the indicator selection criteria.

3.2.1 Selection criteria

There exists tension between convenience and comprehensiveness when selecting an indicator set; a smaller set using easily available data is more convenient to collect and to analyse but may overlook important aspects while a larger set can be more comprehensive but have excessive data collection and analysis costs (Litman, 2005). In literature (Bird et al., 2005; Nardo et al., 2005b; National Cooperative Highway Research Program, 2003; Sharpe, 2004; Zietsman et al., 2008) several criteria for selecting good indicators can be found. The criteria suggested in a number of relevant sources are listed in Table 1. Litman (2007) published a paper on indicators for comprehensive and sustainable transportation planning. Ledoux et al. (2005) and Hens et al. (2005) focused on sustainable, respectively environmental health indicators. The criteria recommended for selecting a good set of indicators in these fields are also valuable for the road safety indicators context. The criteria listed in the two final columns are deduced from road safety studies in which some conditions and requirements for indicators are described (Farchi et al., 2006; Al Haji, 2005).

Table 1: Indicator selection criteria used in literature

Litman (2007)	**Ledoux et al. (2005)**	**Hens et al. (2005)**	**Farchi et al. (2006)**	**Al Haji (2005)**
Understandable	Interpretable	Simple	Interpretable	Understandable
Useful	Comparable	Policy relevant	Clear definition	Clear definition
Comparable	Measurable	Available data	Relevant	Relevant
Available data	Continuous	Valid	Comparable	Comparable
Diverse	Cost effective	Reliable	Theoretically valid	Measurable
Target	Sensitive	Sensitive	Available data	Continuous
	Robust	Specific	Continuous	Reliable
			Timeliness	Accepted
			Reliable	Target
			Stable	
			Cost effective	
			Sensitive	

Table 1 shows that some selection criteria are mentioned in almost all studies while others appear only once. In theory, an indicator should meet all these conditions. However, in practice, it is more interesting to evaluate indicators on a smaller number of essential selection criteria. Based on Table 1 and our own insights, good road safety performance indicators should satisfy eight criteria, i.e., relevant/valid; measurable, understandable, available data, reliable, comparable/coherent, specific and sensitive. For each criterion, a description is given followed by an example of a good safety performance indicator respecting the criterion (+) and a bad indicator falling short of the criterion (-).

1. <u>Relevant/valid</u>: can the indicator be associated with policy objectives? Is the indicator action-oriented and suitable for establishing a performance target? Does the indicator provide a good picture of the phenomenon that we want to measure (i.e., the performance on a specific road safety risk domain)

   +: the percentage of drivers exceeding the speed limit on various road types

-: the intersection density (i.e., the ratio of the number of intersections and the area of a country)[8]

2. Measurable: is the indicator quantifiable and measurable?

   +: the number of hospital beds per 10,000 citizens

   -: the attitude towards drinking and driving

3. Understandable: is the indicator clearly defined? Does the indicator have a comprehensible and acceptable interpretation?

   +: the percentage of motorcyclists wearing a helmet

   -: the median of the set of absolute differences between each of the observed speeds in the road class and the median of all observed speeds divided by the median of the set of observed speeds

4. Available data: are data available for a large set of subjects (countries in this case) within an acceptable term and at a reasonable cost? Can the indicator be updated on a regular basis?

   +: the percentage of surveyed (car) drivers disrespecting the alcohol limit

   -: the percentage of road length with facilities for separation of slow vulnerable traffic and other motorized traffic

5. Reliable: do the data come from a reliable source? Have the data been collected in a scientific way?

   +: the number of emergency medical services' staff per 10,000 citizens

   -: the percentage of persons <12 years correctly sitting in a child's seat in the front or rear seat of a car

6. Comparable/coherent: is the indicator coherent over time (i.e., has the same definition, method, … been used) and over space (i.e., do the subjects apply the same definition, …)?

   +: the motorway density (i.e., the ratio of the motorway network length and the area of a country)

[8] Some indicators suggested in literature are state indicators or effect indicators (see Section 1.5.1) rather than performance indicators and consequently not action-oriented.

-: the percentage of tested road users with an alcohol concentration above the legal limit

7. Specific: does the indicator focus on a certain level? Is the indicator detailed enough?

   +: the average speed per road type and vehicle type at night

   -: the number of hospital beds per 10,000 citizens

8. Sensitive: is the indicator capable of reflecting changes in risk over time? Do small changes manifest in another indicator value?

   +: the percentage of cars failing the official vehicle inspection

   -: the network density (i.e., the ratio of the total network length and the area of a country)

3.2.2 Indicator card

Choosing one indicator over another for monitoring a specific risk factor has major implications, for example different policy measures could be suggested. Therefore, it is advisable to make an overview of the main characteristics of each possible indicator based on which the best indicators for each risk domain can be identified.

The indicator card contains ten aspects which extensively describe the indicator and help in assessing its relative appropriateness. The indicator card shown in Figure 11 results from some changes made to the model presented in Hens et al. (2005). An elaborated example is shown in Appendix I for the indicator 'the percentage of persons wearing a seat belt in the front seats of a car or van'.

The first section gives some basic information about the indicator. More specifically, the risk domain which is measured by this indicator is given as well as a short definition and the measurement unit. Secondly, the indicator is framed in a broader entirety, namely it is indicated to which component of the causal road safety framework (Section 1.5.1 page 18) the indicator belongs and with which other factors it has a link. Policy relevance is the subject of the third section. The degree of significance for policymakers is assessed, thereby taking into account target setting and prevailing regulations. Methodological notes are discussed next. In particular, it is described how the indicator is computed and which data are needed for this. The current measurement method (e.g., survey or observations) and the sampling design, collection frequency and data source are given. Possible limitations and recommended alternatives are also listed.

1. Name Safety Performance Indicator
   1.1) Domain
   1.2) Description and context
   1.3) Measurement unit
2. Position in the general framework
3. Policy relevance
   3.1) Applicability or relevance
   3.2) Link with other indicators
   3.3) Objective and values to aim at
   3.4) Regulation (national, international)
4. Methodological description
   4.1) Measurement method
   4.2) Data needed, collection frequency, source
   4.3) Limitations
   4.4) Alternatives
5. Assessment of the Safety Performance Indicator
   5.1) Relevant/valid
   5.2) Measurable
   5.3) Understandable
   5.4) Available data
   5.5) Reliable
   5.6) Comparable/coherent
   5.7) Specific
   5.8) Sensitive
6. Evolution of the Safety Performance Indicator
7. Comparison between subjects
8. Positive and negative aspects
   8.1) Strengths
   8.2) Weaknesses
9. Conclusion
10. References

Figure 11: Indicator card

Next, section 5 discusses the degree to which the indicator satisfies each of the eight criteria identified in Section 3.2.1 (criteria 5.1 is linked to section 3 of the indicator card). If data are available, the indicator performance can be shown over time and with respect to different countries. In case of frequent measurements each new value can be plotted against previous values. A graph showing all values in time as well as warning lines indicating unsatisfactory performance (i.e., deteriorating performance or failure to reach targets) is valuable in this respect (Bird et al., 2005). On the contrary, a graph showing only a few data points (e.g., this year's and last year's result) may be misleading for reasons including regression to the mean. If possible, one should compare with a recent average level or trend. Based on the information in the first seven sections, the main strengths and weaknesses of the indicator can be identified, supporting the conclusion whether or not the indicator is suitable for monitoring performance in a certain risk domain. Finally, references related to the indicator are listed.

For an indicator with a number of negative aspects (for example not specific enough and unreliable data) its weaknesses, listed under 8.2, can be seen as points to improve. Moreover, as these strengths and weaknesses may vary in time, up-to-date information is required in the indicator card. Indicators characterized by a positive assessment on the selection criteria should be considered as best available indicators. In addition, best needed indicators also score well on most aspects, but have a (few) major shortcoming(s), i.e., no data available; no reliable data; or no comparable/coherent data. For studies on the relative safety performance of a large set of countries data might be a big issue. The availability of road safety data in Europe is handled in the next chapter.

3.3 Evaluation of possible indicators

In this section, possible indicators for the seven risk domains are listed. There are many indicators that may play an essential role and give a good indication of the prevalence of the risk factor but for practical purposes (e.g., collection costs) only appropriate indicators should be selected. Therefore, each indicator is evaluated and a set of 'best' indicators deduced.

Possible indicators can be specified by posing the question: 'given the relationship between the risk factor and road safety, by means of which indicators can this risk domain be appropriately measured?'. The potential indicators listed below have been identified by means of relevant literature (e.g., Al Haji, 2005; Begeleidingscomité, 2001; European Transport Safety Council, 2001; Morsink et al., 2007; Vis and Van Gent, 2007a).

For each risk domain, the evaluation of the indicators is discussed. By taking into account the information on the indicator cards and in particular the assessment with respect to the selection criteria, the best[9] indicators can be identified. The availability of reliable and comparable indicator data possibly being a critical aspect, a distinction between **best needed** and best available road safety indicators will be made (European Commission, 2005). An ideal (or best needed) road safety performance indicator set can be formulated, irrespective of whether the required data are available, reliable and comparable over time and space. Judging each possible indicator presented below based on the following five, not data related selection criteria – relevant, measurable, understandable, specific and sensitive – a set of best needed indicators for the risk domains alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management can be deduced. The overall assessment considering all eight selection criteria determines the best available indicators. In Appendix II the detailed evaluation scores of all possible indicators are presented. The results in terms of best indicators are given for each risk domain successively.

3.3.1 Alcohol and drugs indicators

The occurrence of alcohol and drugs in traffic can be measured by the following indicators:

- the percentage of tested road users with an alcohol concentration above the legal limit
- the percentage of tested road users with a drugs concentration above the legal limit
- the percentage of surveyed (car) drivers disrespecting the alcohol limit
- the percentage of road user population impaired by alcohol or drugs
- the percentage of road user population impaired by alcohol and drugs

[9] Best indicators rather than appropriate ones will be combined in the composite road safety performance index. By setting a threshold value appropriate indicators could be distinguished from inappropriate ones. Nevertheless, the aim is to identify the most appropriate or best indicator for representing a particular risk domain.

- the percentage of fatalities resulting from accidents involving at least one driver impaired by alcohol
- the percentage of fatalities resulting from accidents involving at least one driver impaired by drugs
- the percentage of fatal accidents in which someone was drinking and driving
- the percentage of road users involved in fatal accidents impaired by alcohol or drugs
- the subjective risk of getting caught while drinking and driving
- the attitude towards drinking and driving

From Appendix II it can be seen that the most relevant, measurable, understandable, specific and sensitive indicator of the ones listed above is **the percentage of road user population impaired by alcohol or drugs**. Furthermore, using all eight criteria in the evaluation – thus also the availability of data, reliability and comparability – <u>the percentage of surveyed (car) drivers disrespecting the alcohol limit</u> appears to be the best available alcohol and drugs indicator.

3.3.2 Speed indicators

The following indicators can be used in relation to speed:

- the percentage of tested drivers exceeding the speed limit on various road types
- the percentage of surveyed (car) drivers exceeding the speed limit on various road types
- the average speed per road type and vehicle type during daytime
- the average speed per road type and vehicle type at night
- the variation in speed per road type and vehicle type
- the median (or another percentile) of the set of observed speeds divided by the speed limit of the road class
- the median of the set of absolute differences between each of the observed speeds in the road class and the median of all

observed speeds divided by the median of the set of observed speeds

- the percentage of drivers with an inappropriate headway on various road types

Based on the evaluation scores, three best needed speed indicators can be identified, i.e., **the average speed per road type and vehicle type during daytime**; **the average speed per road type and vehicle type at night**; and **the variation in speed per road type and vehicle type**. Furthermore, the best available speed indicator is <u>the percentage of surveyed (car) drivers exceeding the speed limit on various road types</u>.

3.3.3 Protective systems indicators

The protective systems indicators are expressed in terms of wearing rate. The wearing rate is the percentage of persons wearing the protective system divided by the total number of users of a certain vehicle type observed in a representative sample during an independent roadside survey (Vis, 2005). Possible indicators for testing the use of protective systems in traffic are:

- the percentage of persons wearing a seat belt in the front seats of a car or van
- the percentage of persons wearing a seat belt in the rear seats of a car or van
- the percentage of persons under 12 years (correctly) sitting in a child's seat, in the front or rear seat of a car
- the percentage of persons wearing a seat belt in the front seats of a bus (above 3.5 tons) or a truck
- the percentage of persons wearing a seat belt in the passenger seats of a bus
- the percentage of cyclists wearing a helmet
- the percentage of moped riders wearing a helmet
- the percentage of motorcyclists wearing a helmet

<u>The percentage of persons wearing a seat belt in the front seats of a car or van</u> is the indicator which scores best with respect to the eight

selection criteria (relevant, measurable, understandable, available data, reliable, comparable, specific and sensitive). Given the diversity of the various protective systems indicators and their good evaluation results, seven indicators can be pointed out as best needed indicators for the third risk domain: **the percentage of persons wearing a seat belt in the front seats of a car or van**; **the percentage of persons wearing a seat belt in the rear seats of a car or van**; **the percentage of persons under 12 years (correctly) sitting in a child's seat in the front or rear seat of a car**; **the percentage of persons wearing a seat belt in the front seats of a bus (above 3.5 tons) or a truck**; **the percentage of persons wearing a seat belt in the passenger seats of a bus**; **the percentage of moped riders wearing a helmet**; **and the percentage of motorcyclists wearing a helmet**.

3.3.4 Daytime running lights indicators

With respect to the daytime running lights domain the following indicators can be considered:

- the total usage rate of daytime running lights
- the usage rate of daytime running lights per road type and vehicle type
- the existence of a law – fully or partially[10] – obligating the use of daytime running lights

The best needed indicator is **the usage rate of daytime running lights per road type and vehicle type** scoring best with respect to relevance and specificity. In case the availability of data, the reliability and the comparability are also incorporated in the evaluation process, all three indicators appear to score relatively poor. However, the existence of a law – fully or partially – obligating the use of daytime running lights is considered to be the best available daytime running lights indicator.

[10] i.e., on certain road types, for certain vehicle types or during certain time periods.

3.3.5 Vehicle indicators

Unsafe vehicle aspects relate to technical defects, crashworthiness, the level of safety equipment and compatibility. Therefore, the following indicators can be used for the vehicle domain:

- the percentage of cars failing the official vehicle inspection
- the EuroNCAP score: the percentage of new cars obtaining 0 respectively 1, 2, 3, 4 or 5 stars in the total number of new passenger cars
- the age distribution of the vehicle fleet: the percentage of vehicles of maximum 5 years; between 6-10 years, between 11-15 years and older than 15 years in the total number of registered vehicles (focus on passenger cars)
- the composition of the vehicle fleet: the percentage of cars, vans, buses, trucks and motorcycles in the total number of registered vehicles

Notice that each percentage can be considered as an individual indicator: the percentage of new cars with 5 stars or the percentage of vehicles older than 15 years or the percentage of trucks. Combining the indicators within a certain category is another option, e.g., the percentage of new cars obtaining 4 or 5 stars.

The evaluation resulted in one indicator being the best needed and at the same time the best available vehicle indicator, namely **the age distribution of the vehicle fleet: the percentage of vehicles of maximum 5 years; between 6-10 years, between 11-15 years and older than 15 years in the total number of registered vehicles (focus on passenger cars)**.

3.3.6 Roads indicators

The roads domain can be characterized by the following indicators:

- the share of network length per road type (e.g., motorway)
- the share of intersections per type (e.g., roundabout, signalised T-junction)
- the intersection density

- the network density
- the motorway density
- EuroRAP road protection scores
- the percentage of road length with a wide median or median barrier
- the percentage of road length with a wide obstacle-free zone or roadside barrier
- the percentage of road length with facilities for separation of slow vulnerable traffic and other motorized traffic
- the percentage of the road network satisfying the safety design standard
- the expenditure on roads (in terms of engineering or maintenance) as share of the gross domestic product (GDP).

Two best needed roads indicators were identified, i.e., **the percentage of road length with a wide median or median barrier** and **the percentage of road length with a wide obstacle-free zone or roadside barrier**. Including the assessment of data availability, reliability and comparability pointed out <u>the motorway density</u> as the best available indicator at this moment.

3.3.7 Trauma management indicators

Possible indicators for the trauma management domain are:

- the percentage of calls to emergency medical services (EMS) due to a road accident
- the average arrival time of emergency medical services at the accident scene
- the number of emergency medical services' staff per 10,000 citizens
- the number of hospital beds per 10,000 citizens
- the average length of stay in the hospital after a road accident
- the share of road casualties treated in intensive care units

- the share of road casualties who died during hospitalisation
- the expenditure on health care as share of the gross domestic product

The average arrival time of emergency medical services at the accident scene and **the share of road casualties who died during hospitalisation** appear to be the best needed trauma management indicators. The best available indicator is <u>the expenditure on health care as share of the gross domestic product</u>.

In the next section, some conclusions in terms of the selection of best available and best needed road safety performance indicators are given.

3.4 Concluding remarks

In this chapter, the first step in the creation of a road safety performance index, i.e., the selection of road safety performance indicators, has been handled. The European Transport Safety Council advised the European Union in 2001 to formulate and specify a set of relevant safety performance indicators that can be used on the European and national level as a means to determine trends in the level of safety and the success of casualty reduction programmes. Starting from seven main road safety risk domains – alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management (Vis, 2005) – we aim for a set of safety performance indicators suitable for capturing road safety risk in a country.

Several indicators exist for each risk domain. To select appropriate indicators the pros and cons of various indicators need to be weighed against each other. All relevant information concerning each indicator can be summarized on an indicator card, in which among other things a definition, some methodological aspects and the degree of meeting the required selection criteria are mentioned. In the evaluation of safety performance indicators eight criteria are in our opinion essential, i.e., a safety performance indicator should be relevant, measurable, understandable, have data available, be reliable, comparable, specific and sensitive.

The availability of reliable and comparable data has a major influence on the indicator selection. To overcome the current, partial lack of indicator information and to assure the creation and collection of necessary data for road safety policymaking in a longer time perspective, best available as well as best needed road safety indicators are identified (European Commission, 2005). An ideal road safety performance indicator set can be formulated, irrespective of whether the required data are available, reliable and comparable over time and space. Judging each possible indicator presented in Section 3.3 based on five, not data related selection criteria – relevant, measurable, understandable, specific and sensitive – resulted in the following set of **best needed indicators** for the risk domains alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management:

- the percentage of road user population impaired by alcohol or drugs
- the average speed per road type and vehicle type during daytime and at night
- the variation in speed per road type and vehicle type
- the seat belt wearing rate in front and rear seats of cars and vans, busses and trucks
- the (correct) usage rate of child's seats in cars
- the helmet wearing rate of motorcyclists and moped riders
- the usage rate of daytime running lights per road type and vehicle type
- the age distribution of the vehicle fleet
- the percentage of road length with a wide median or median barrier
- the percentage of road length with a wide obstacle-free zone or roadside barrier
- the average arrival time of emergency medical services at the accident scene
- the share of road casualties who died during hospitalisation

As data issues currently limit the use of best needed indicators, clear data collection recommendations are needed. In view of a coherent and comparable indicator data set, best practices in terms of data collection need to be advocated. In this respect, one could benefit from the manual developed at the European level, specifying the measuring of indicators and sampling designs (Hakkert and Gitelman, 2007). In the meantime, the road safety index will consist of best available indicators characterized by available data of acceptable quality. Here, the index methodology will be illustrated using real road safety performance indicator data (more information on the data is given in Chapter 4). More generally, the following indicator set (scoring best on all eight selection criteria) can be considered to be the best available one at this moment:

- the percentage of surveyed car drivers disrespecting the alcohol limit
- the percentage of surveyed car drivers exceeding the speed limit on various road types
- the seat belt wearing rate in front seats of a car or van
- the existence of a law – fully or partially – obligating the use of daytime running lights
- the share of the fleet of passenger cars per age class
- the motorway density
- the expenditure on health care as share of the gross domestic product

The above list shows that the unavailability of good data restricts the selection of the indicator set to some extent. Self-reported data are to be used for the alcohol and speed domain and for the roads domain there are no direct or indirect safety performance indicators in use in Europe at the moment (Hakkert et al., 2007). Moreover, the evaluation of the daytime running lights and trauma management indicators revealed that proxies are to be used for now.

Given the difficulties encountered in the identification of good road safety performance indicators and in order to avoid unnecessary complexities in the illustration of the multistage index process, only one – best available – indicator is used to represent each of the seven risk domains at this point. Nevertheless, the index methodology presented in this manuscript is still valid in case several best or appropriate indicators per domain would be available. The methodological process for combining indicators needs to be applied more than once then (see Figure 12). First, the methodology is used on the set of alcohol and drugs indicators to create an alcohol and drugs index. The same applies to the other six domains. Finally, the alcohol and drugs index, the speed index, the protective systems index, the daytime running lights index, the vehicle index, the roads index and the trauma management index are joined in one overall road safety performance index. So in theory, the road safety performance index consists of seven domain indexes, each one being an aggregation of specific individual indicators. The index methodology is illustrated focusing on the final combination process.

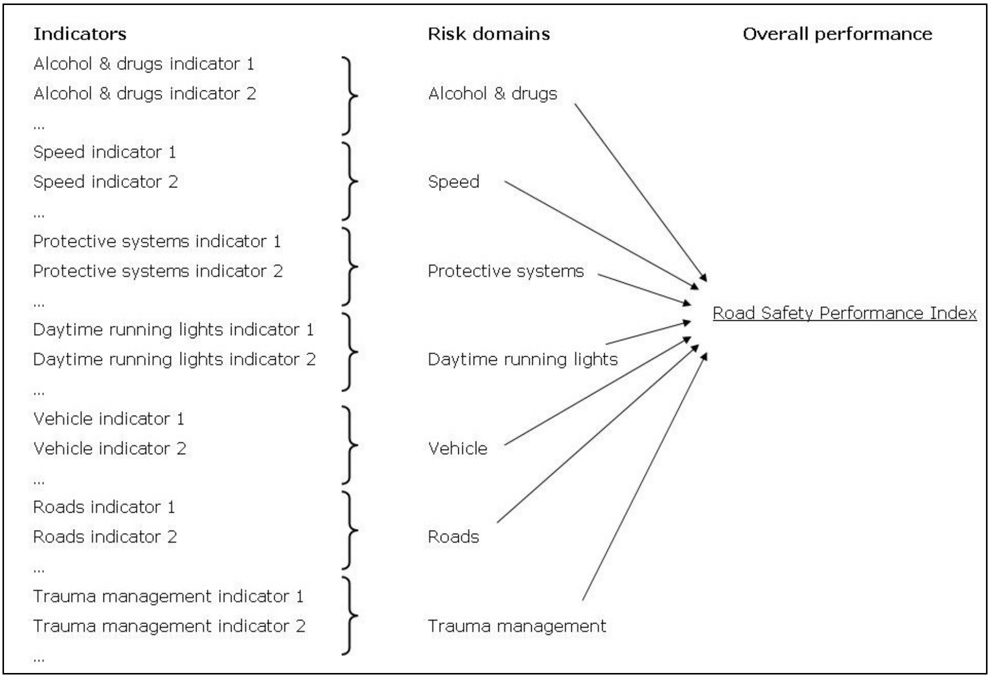


Figure 12: Structure of the road safety performance index

To conclude, it is important to note that the selection of appropriate safety performance indicators is a process requiring periodic revisions. The set of best needed and best available indicators needs updating and refinement from time to time. Depending on the risk domain, the indicators call for different kinds of development efforts relating to concepts, methodologies and data collection procedures (European Commission, 2005). The indicators should be produced with adequate frequency to support decision makers in a systematic way, i.e., for identifying problems, monitoring implementation of measures and following-up their effectiveness. The next chapter discusses the availability and quality of road safety data in Europe and presents the indicator values used in the creation of the composite index scores.

Chapter 4 Road safety data

Quantification is at the core of scientific understanding and requires data (Evans, 2004). Data on the incidence and types of accidents as well as a detailed understanding of the circumstances that lead to accidents is required to guide safety policy (World Health Organization, 2004). In this chapter, existing road safety data sources are reviewed and data are collected. More specifically, Section 4.2 gives an overview of relevant data sources with respect to each of the seven risk domains. The best available indicator values that will be used in the creation of the index are discussed. Apart from safety performance indicator data, a number of final outcome data sources are described (Section 4.3). Concluding remarks with respect to the data can be found in Section 4.4. First, some general remarks are given in Section 4.1.

4.1 Introduction

Accident data are the most common type of road safety data. Nearly all countries in the world record accidents. Since the beginning of motorization, detailed information (i.e., the number of accidents, the number of casualties and their characteristics) has been gathered and inventoried on a regular basis. Despite this large amount of data, many questions remain unanswered because other factors of interest have not been recorded or the data are not sufficiently reliable, complete or conveniently accessible (Evans, 2004).

Data quality is affected by a number of factors such as the sources of data, underreporting and definitional issues. There are several sources of accident data: police reports, hospital records, insurance company information and accident involvement surveys. Even though hospital admission statistics, insurance statistics and accident involvement surveys result in a higher level of reporting of injury accidents, they suffer from a lack of information about the location of the road accidents and of other parties involved in the accidents (see e.g., Christens, 2003). Therefore, the official police data are generally used in road safety research. Farchi et al. (2006) revealed only small differences between the countries of the European Union concerning the type of accident data collected. The most commonly reported characteristics are related to the accident (location, time of occurrence, light conditions), the persons involved (age, sex, physical condition, use of safety devices, type of road user) and the vehicles involved (type). These accident data are sometimes linked to census data and exposure data to draw conclusions in terms of mortality rate, fatality rate and fatality risk.

However, for most countries, we may conclude that the reporting of accidents in official statistics is inaccurate and incomplete (Elvik and Vaa, 2004). Tessmer (1999) showed there is a problem of under-recording by the police in most countries. Some detailed information of the accident is lacking as policemen have limited medical or engineering background. Additionally, a number of accidents are never reported to official authorities. In general, the level of reporting increases with accident severity. Furthermore, accidents involving children or bicyclists, single-vehicle accidents and accidents in rural areas have a low police reporting level (Christens, 2003). Inconsistency in the level of reporting on a disaggregated (for specific types of accidents) as well as an aggregated level (across countries) as well as under-recording are important data shortcomings that should be kept in mind.

Differences in injury categorization of accidents across Europe is a next aspect. In order to fairly compare European countries with respect to road safety the definition of important concepts should be the same. In recent years, efforts have been made in this respect. Nowadays, the number of persons killed within 30 days after an accident is typically used in international databases. The European Commission 6th Framework integrated SafetyNet project aims to accelerate the availability and use of harmonised road safety data in Europe. One of the main outputs of this project is the European road safety observatory (ERSO) website (www.erso.eu) which is a system designed for road safety professionals to bring together road safety related data and knowledge and provide access to this information. In the SafetyNet work package on safety performance indicators, which is of most interest for our research, the availability of safety performance data was investigated in 27 cooperating countries (25 member states plus Norway and Switzerland) (Vis and Van Gent, 2007b). It is stated that in general, it is difficult to compare the road safety performance of countries (Vis and Van Gent, 2007a). The main reasons are lack of data (essential data are missing), suspicious quality of the data received from numerous authorities (affecting the validity of the results) or the incomparability of (seemingly similar) data due to different circumstances of measurement. Moreover, some EU countries have laws prohibiting the measurement of necessary data (e.g., alcohol and drugs substances present in the road user population). It can be concluded that there is still a lot of work to be done before the availability of high-quality and comparable safety performance indicator data will be a fact. In this respect, the development of a manual (Hakkert and Gitelman, 2007) providing details on the procedures for gathering the required indicator data contributes to uniform data collection in Europe.

Accurate safety performance data enable a country to diagnose its road safety problems and to select appropriate measures to apply. The availability of reliable and comparable indicator values is considered in the evaluation process for selecting best available indicators. Although the use of imperfect (or no ideal) indicators influences the outcomes to some extent they are – at least for now – needed to create a workable indicator data set for illustrating the index methodology. Recent indicator values are gathered for as many European countries as possible. Since we aim to compute an overall index score for each country, the indicator data in the various sources should relate to the same set of countries and the same period (or year).

4.2 Data sources on safety performance indicators

In the previous chapter, seven risk domains were described. For each domain, possible indicators were listed and subsequently evaluated with respect to relevance, measurability, interpretability, data availability, reliability, comparability, specificity and sensitivity (see Appendix II). Based on the overall score, best available indicators were identified. In the current chapter, the availability of reliable and comparable indicator data is studied in more detail. More specifically, data sources for the best available indicators will be searched. In addition, other sources in which valuable data for each risk domain can be found will be indicated. The best available indicator values that will be used in the creation of the index are discussed.

4.2.1 Alcohol and drugs

4.2.1.1 Relevant sources

In the past, more research has been devoted to alcohol than to drugs. For this first risk domain we therefore focus on alcohol. Alcohol data can be found in several sources. Some important ones are listed below:

- The Social Attitudes to Road Traffic Risk in Europe (SARTRE) research: the percentage of car drivers reporting to have been driving with a blood alcohol content above the legal limit at least one day during the past week. Data relating to 2002-2003 are available for 23 countries.
- The United Nations Database: the percentage of accidents or fatalities in an accident with at least one person drinking and driving. 2002 data are available for 20 countries.
- The alcohol database of the World Health Organization: the number of road accidents in which alcohol was involved per 100,000 accidents. The year to which the data relate strongly differs between the countries.
- The SafetyNet project: the percentage of fatalities resulting from accidents involving at least one driver impaired by alcohol. Data relating to the time frame 2001-2005 are available for 21 countries.

Country-specific sources such as reports from national road safety institutes may provide additional information. Nevertheless, comparable values for a large set of countries are not guaranteed in that case (e.g., due to a different measurement method). However, these figures can be used to compare to or to complete some missing information.

The best available alcohol and drugs indicator appeared to be the percentage of surveyed car drivers disrespecting the alcohol limit. The first data source listed above is therefore chosen to provide data for the alcohol and drugs domain. Compared to the other data sources, values with respect to an indicator defined on the level of intermediate outcomes (see Figure 5) are given whereas the indicators of the other three sources refer to the level of final outcomes and hence are less relevant for a road safety performance index. We opt for the first source providing values for a large set of countries referring to the same time period. More information on the indicator values is given in the next section.

4.2.1.2 SARTRE

The Social Attitudes to Road Traffic Risk in Europe research is a survey based research that has been conducted three times in several European countries. Here, we focus on the most recent SARTRE research (SARTRE 3 consortium, 2004). The national surveys were launched between September 2002 and April 2003 in 23 countries. Fourteen of the EU-15 countries were involved (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and United Kingdom; Luxembourg being the only one not taking part), seven at that time 'applicant' countries (Cyprus, Czech Republic, Estonia, Hungary, Poland, Slovakia and Slovenia) as well as Switzerland and Croatia. Within each country, a representative sample of active car drivers was selected. The sampling method took care of geographical distribution and rural-urban balance and respondents were chosen at random or based on quota.

The SARTRE 3 survey gathered and analysed data on several road safety aspects in 23 countries. In relation to alcohol, the following topics were discussed:

- drinking behaviour and habits
- opinions about alcohol risk when driving
- opinions about alcohol legislation
- reported experiences and opinions about enforcement
- opinions about measures to prevent drink driving

As safety performance indicator for the alcohol and drugs domain we select the percentage of car drivers reporting to have been driving with a blood alcohol content above the legal limit (at least one day during the past week). The data are shown in Figure 13. It should be noted that in case of self-reporting, data might be biased to some extent. However, these data are considered to be the best available alcohol performance values and will be used in the creation of the road safety index.

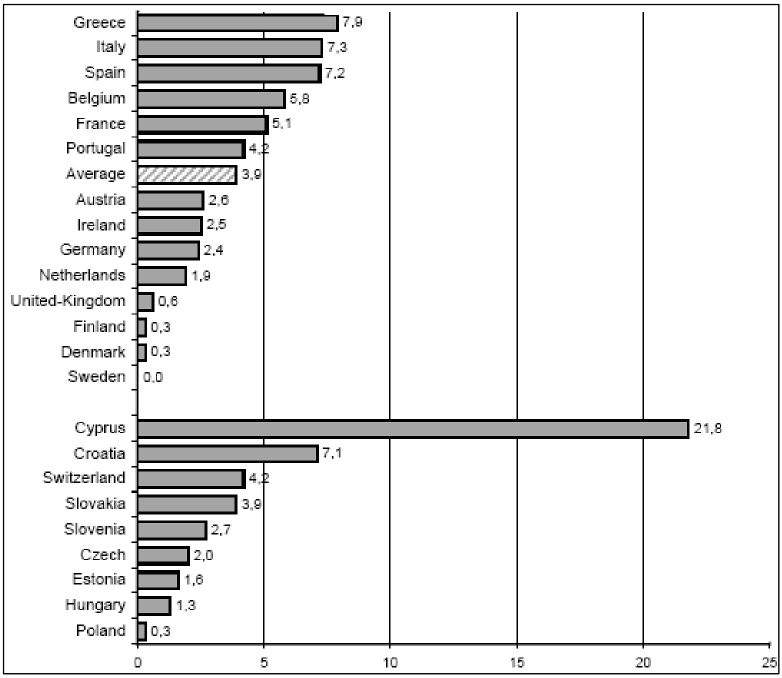


Figure 13: Percentage of respondents reporting to have driven at least one day in the past week when they may have been over the legal limit for drinking and driving (SARTRE, 2004)

4.2.2 Speed

4.2.2.1 Relevant sources

For the speed domain a similar approach is followed. First, some sources for speed data are given. They relate to speed violations on the one hand and average speed on the other.

- The Social Attitudes to Road Traffic Risk in Europe research: the percentage of car drivers reporting to often, very often or always exceed the speed limit on motorways, main roads between towns, country roads and built-up areas. Data relating to 2002-2003 are available for 23 countries.
- The European Transport Safety Council: the percentage of drivers exceeding the speed limit on urban roads, rural roads and motorways. Data relating to 2005 are available for 11 respectively 15 and 13 countries (depending on the road type).

- The SafetyNet project: the percentage of cars (all traffic together for one country) exceeding the speed limit on motorways. Data are available for 5 countries for several years within the period 1999-2005.
- The European Transport Safety Council: the average speed on urban roads, rural roads and motorways. Data relating to 2005 are available for 12 respectively 16 and 14 countries (depending on the road type).
- The SafetyNet project: the average speed of cars (all traffic together for two countries) on motorways. Data are available for 6 countries for several years within the period 1999-2005.

Again, country-specific sources may provide additional information but usually not end up in comparable values for a large set of countries.

The previous chapter concluded that the percentage of surveyed car drivers exceeding the speed limit on various road types is the best available speed indicator. Moreover, the first data source has the largest country coverage. An advantage of using the same data source for both the alcohol and drugs domain and the speed domain is that the same set of countries and the same time period are considered for both indicators. More information on the indicator values is given in the next section.

4.2.2.2 SARTRE

With respect to speeding the SARTRE 3 survey collected information about:

- driver's perceptions of their speed and safety compared to other drivers
- speed behaviour such as exceeding the speed limit, driving through amber
- opinions about the contribution of speeding to accidents
- existing speed limits and enforcement activity
- opinions about their enjoyment of speed
- accident involvement

The percentage of car drivers reporting to often, very often or always exceed the speed limit on a certain road type is the chosen speed indicator. As shown in Table 2, the frequency of speed limit exceeding depends on the road type. It can be seen that it is the highest on motorways and the lowest in built-up areas.

Table 2: Percentage of respondents reporting to exceed the speed limit either often, very often or always, on various road types (SARTRE, 2004)

	MOTORWAY	MAIN ROADS BETWEEN TOWNS	COUNTRY ROADS	BUILT-UP AREAS
Austria	19	11	11	6
Belgium	27	17	14	12
Denmark	46	34	14	4
Finland	17	11	10	6
France	22	14	10	7
Germany	20	15	17	7
Greece	40	23	19	6
Ireland	10	7	4	3
Italy	24	26	15	12
Netherlands	31	22	14	7
Portugal	32	19	15	11
Spain	37	21	13	11
Sweden	35	27	14	5
Unt Kingdom	26	13	8	4
Average	*28*	*19*	*13*	*7*
Croatia	25	18	21	6
Cyprus	28	21	18	12
Czech Rep	14	12	7	6
Estonia	13	25	20	12
Hungary	16	21	17	12
Poland	12	13	11	7
Slovakia	16	18	11	8
Slovenia	26	16	10	6
Switzerland	32	21	18	4

In this research, only one indicator will be used to represent each risk domain (see Section 3.4). Therefore, we select one of the four road types rather than using the average of the four values for each country as this might mitigate the differences between countries. In further research, it would be useful to construct a composite speed index in which the values of each road type are weighted by the share of kilometres driven on that road type in the country. However, a different, more generally agreed upon road classification (see e.g., the European Union Road Federation or the SafetyNet project) might be more appropriate. For now, we make a selection based on the correlation of each speed indicator with the chosen final outcome indicator (see Section 4.3)[11]. Consequently, the indicator for the speed domain is the percentage of car drivers reporting to often, very often or always exceed the speed limit in built-up areas.

4.2.3 Protective systems

4.2.3.1 Relevant sources

For the protective systems domain, indicators with respect to seat belts, child's seats and helmets can be defined. Available data mainly relate to the seat belt wearing rate in the front seats. European values for the best available indicator, the seat belt wearing rate of occupants in the front seats of a car or van, can be found in the following publications:

- European Transport Safety Council's 'Seatbelt wearing in the EU: Fact sheet April 2006'. 2004 data are available for 16 countries and 2003 data for 4 countries. In a related document (available on http://www.etsc.be/enforcement-seatbeltuse-whyincrease.php) the same values are given. However, here it is also indicated whether the data apply to drivers only or both drivers and passengers in front.
- European Transport Safety Council's 'PIN flash 4 on seat belt use (February 2007)' in cooperation with SafetyNet. The most recent wearing rates are given. For most countries the value refers to

[11] The road safety performance index that we are building – situated on the level of intermediate outcomes – should have a clear link with the final outcome level. That way, the index and indicators can act as warning signals and specific measures can be taken before the risk becomes visible in an increased number of casualties.

2005, for some to 2006, for one country to 2002 and for another one to 2003. Driver wearing rates are available for 20 countries; overall wearing rates in the front seats for 24 countries. They differ three percentage points at most. Moreover, explanatory notes are given (e.g., wearing rates for cars only; on urban and rural roads only; ...) and should be kept in mind when comparing values across countries. The data in the table of PIN flash 4 are also graphically presented in Vis and Van Gent (2007a).

4.2.3.2 ETSC

Data for the seat belt wearing rate in the front seats of cars and vans are deduced from the first source as the time period 2003-2004 corresponds best to the 2002-2003 SARTRE data obtained so far. Values are available for 20 countries (see Table 3). Although we may expect that there is a change in seat belt wearing behaviour over the years, we assume for now that the values for 2003 and 2004 are comparable. The same assumption is made for the wearing rate involving drivers only and the rate for all occupants in the front seats.

Comparing the set of countries with a seat belt value to the 23 countries participating in the SARTRE research reveals that for five countries – Croatia, Cyprus, Italy, Slovakia and Switzerland – no seat belt value is available in Table 3. Therefore, the second source, i.e., 'PIN flash 4 on seat belt use' is consulted for these countries. In fact, the seat belt wearing rate in front seats for Cyprus is 80 (this most recent value relates to 2002); for Italy 71 (this value relates to 2005 and does not include motorways); and for Switzerland 82 (representing the seat belt wearing rate of drivers in 2005). Despite 2002 and 2005 values we prefer to include as many countries as possible (i.e., 21 countries; Croatia and Slovakia no longer being considered) in order to make a fair road safety performance comparison.

Table 3: Seat belt wearing rate in front seats (ETSC, 2006)

	WEARING RATE FRONT SEATS (%)	REMARK
Austria	77	2004
Belgium	66	2004
Cyprus	n/a	
Czech Rep	75	driver 2004
Denmark	84	2004
Estonia	75	2004
Finland	89	2004
France	97	2004
Germany	94	driver 2004
Greece	40	2003
Hungary	59	2004
Ireland	85	2003
Italy	n/a	
Latvia	n/a	
Lithuania	n/a	
Luxembourg	88	driver 2004
Malta	95	driver 2004
Netherlands	86	2003
Poland	71	2004
Portugal	88	driver 2004
Slovakia	n/a	
Slovenia	81	2004
Spain	86	2003
Sweden	92	driver 2004
Unt Kingdom	93	driver 2004

4.2.4 Daytime running lights

4.2.4.1 Relevant sources

Sources providing relevant data with respect to the daytime running lights domain are rather sparse.

- The SafetyNet project: daytime running lights usage rates on different road types (motorways, rural roads, urban roads and DRL roads). Data are available for 8 countries and relate to different years (1993, 2002, 2004, 2005 and 2006).
- The consultation paper on DRL (European Commission, 2006b): legal obligation of daytime running lights in European countries (per road type and time of the year). The paper indicates 14 European countries in which the use of daytime running lights is mandatory to some extent. DRL applies to all roads throughout the whole year in Austria, Czech Republic, Denmark, Estonia, Finland, Latvia, Slovenia and Sweden. A partial law exists in Hungary, Italy, Lithuania, Poland, Portugal and Slovakia.

Both sources are evaluated below.

4.2.4.2 Discussion

The first source lists DRL usage rates for a limited number of countries. In Northern countries the rates are not regularly measured as due to a long history of DRL and a large share of automatic switch-on lights the overall rate is close to 100% (Vis and Van Gent, 2007a). In other countries, this aspect receives rather limited attention thereby yielding data unavailability. Moreover, the difference in year to which the SafetyNet figures relate might be influential. Therefore, the second source seems to be the more appropriate one in terms of data availability.

However, choosing this DRL law indicator (the best available indicator for the daytime running lights domain) implies a categorical rather than a continuous indicator. More specifically, three classes may be distinguished: countries obligating the use of DRL on all roads during the entire year; countries obligating DRL on some roads and/or during some periods of the year; and countries enacting no DRL law (possibly recommending it). As the second source gives information on the first two classes, the countries not listed there are considered to belong to the third class. However, we cannot be entirely certain about the classification and the resulting indicator values. Within the country comparisons deliverable of SafetyNet (Vis and Van Gent, 2007a) apart from some DRL usage rates a brief description of the DRL law is given for a limited set of countries. From that information it can be deduced that some countries only recently enacted a DRL law, being Austria since 2006 and the Czech Republic since 2004. As the indicator data set used in this manuscript focuses on 2003, this should be taken into account.

To conclude, the best available daytime running lights indicator already scored below average at the evaluation (in Appendix II) and appears to be problematic. The performance of countries with respect to daytime running lights is not captured by this indicator, the nature of the indicator distinguishes only three possible values and the classification is characterized by some level of uncertainty. Therefore, it is decided to incorporate no indicator for the daytime running lights domain in the index construction process at this moment.

4.2.5 Vehicle

4.2.5.1 Relevant sources

For the fifth risk domain, four types of indicators were deduced from literature: vehicle inspection, EuroNCAP, age distribution and composition of the vehicle fleet. Vehicle related indicator data can be found in several sources. Below, the main sources are given:

- United Nations Economic Commission for Europe as well as Eurostat: passenger cars per age group (0-2 years; 3-5 years; 6-10 years; >10 years) at 31/12/2002. Both sources mention the same values, except that in the latter source (Eurostat) 2002 data for Finland are given as well, so in total, values are available for 19 countries.
- The SafetyNet project: passenger cars per age group (<1994; 1994-1998; 1999-2003; >2003). 2003 data are available for 15 countries.
- United Nations Economic Commission for Europe: road vehicle fleet composition per vehicle type (e.g., moped; motorcycle; passenger car; bus; lorry; etc) at 31/12/2002. Data are available for a large set of countries (although with some missing values).

- The SafetyNet project: road vehicle fleet composition per vehicle type (car/taxi; lorry<3.5 tons; heavy goods vehicle; agricultural tractor; bus; motorcycle). Data are available for 11 countries.
- The SafetyNet project: percentage of the road vehicle fleet tested by EuroNCAP. Data with respect to 2003 are available for 11 countries.

Apart from these sources, data with respect to vehicle inspections are collected in most countries. However, the coherence of these values might be a problem.

The evaluation of possible vehicle indicators concluded that the share of the fleet of passenger cars per age class is the most appropriate indicator to measure the vehicle domain at this moment. Three sources (UNECE, Eurostat and SafetyNet) provide data for this indicator. In order to compare the safety performance of as many countries as possible, we will start with the data from Eurostat.

4.2.5.2 Eurostat

Table 4 shows the total number of registered passenger cars by the end of 2002 for 22 countries. In addition, the number of cars in each of the four defined age classes is available for 19 countries. As indicator for the vehicle domain, we opt for the percentage of passenger cars of maximum five years old. This share – not too narrow and relevant in terms of technological improvements – is expected to be a good indication of the safety level present in the vehicle fleet. By dividing the sum of the values in columns 3 and 4 by the total number of passenger cars an indicator value can be obtained for 19 countries (e.g., 40.84% for Belgium).

Table 4: Number of passenger cars per age class (Eurostat, 2008)

	TOTAL	<2 Y	2 – 5 Y	6 – 10 Y	>10 Y
Belgium	4787	726	1229	1420	1412
Czech Rep	3647.1	261.3	447.3	808.5	2129.9
Denmark	1888	306	433	546	603
Germany	44657	6597	10307	14431	13322
Estonia	400.7	26.7	41.9	53.4	278.7
Ireland	1470	289	500	463	197
Greece	3646				
Spain	18733	2758	4007	4418	7550
France	29160	4553	6288	8915	9404
Italy	33706	4575.229	7309	8699	13123
Cyprus	288	18	31	112	127
Hungary	2629.5	309	382	454.1	1484.4
Netherlands	6855	1012	1655	2045	2143
Austria	3987	535	866	1296	1290
Poland	11028.9	734.5	1724.2	2533.7	6036.5
Portugal	5788				
Slovenia	914.5	227.8	230.1	274.8	181.9
Slovakia	1326.9				
Finland	2195	228	404	409	1138
Sweden	4044	431	1047	798	1472
Unt Kingdom	26460	5034	6419	8493	5837
Switzerland	3701	599	867	1131	1103

Values for the alcohol and drugs, speed and protective systems indicator were obtained for 21 European countries. To match this set of countries the percentage of 'young' (i.e., under six years) passenger cars or an approximated value needs to be found for Portugal and Greece as well. Using the Eurostat database (Eurostat, 2008) the most recent detailed values for Portugal appear to relate to 2000. (673+836)/5260 equals 28.7%. Assuming that this share remained more or less stable within the 2001-2002 period, this value will be used in our data set. For Greece, a proxy value can be obtained via the second source in 4.2.5.1. From the SafetyNet classification of passenger cars it appears that the share of recent cars in Greece is similar to the one in Sweden, Spain, Denmark and Austria. Therefore, the average of these four values (based on Table 4) is taken to approximate the vehicle indicator value for Greece.

4.2.6 Roads

4.2.6.1 Relevant sources

A high number of possible indicators for the roads domain were found in literature. However, data issues limit the selection of an appropriate performance indicator. The following sources provide information with respect to roads:

- The European Union Road Federation: share of the road network by category (i.e., motorways; national roads; secondary or regional roads; and other roads[12]). 2003 values are given for 32 countries.
- The International Road Traffic and Accident Database: network density. For 13 countries in the set of 21 countries considered so far values are available with respect to 2003. Four countries have less recent values. In addition, other sources (e.g., United Nations) provide data on the length of the road network and the area for a large set of countries.
- The International Road Traffic and Accident Database: motorway density. For 13 of the 21 countries considered so far data are available with respect to 2003. Four countries have less recent values. Additionally, the Eurostat yearbook 2006-2007 provides motorway density values for 20 of the 21 considered countries with respect to 2003. The European Union Road Federation graphically presents the density of motorways in 32 countries in a particular year.
- The SafetyNet project: intersection density and the share of roads with a wide median or median barrier (for different types). Data are available for 8 countries.

[12] The definition of 'other roads' is not harmonised across countries.

- The European Union Road Federation: investment in road maintenance per kilometre of road network[13]. Data are available for 18 countries (with respect to different years, i.e., 1999, 2003, 2004, 2005 and 2006).

From the evaluation of possible indicators it was concluded that the density of motorways is currently the best available indicator for the roads domain. Contrary to the network density, the relationship between motorway density and road safety is much clearer because motorways are considered to be the most safe road type (in terms of fatalities per kilometre travelled). Additionally, a high density of motorways implies that persons are rather close to an approach road which enables them to use this safest road type more quickly. The data used are presented in the next section.

4.2.6.2 Eurostat

There are three sources (IRTAD, Eurostat and ERF) providing motorway density data. Here, we select the Eurostat yearbook because it presents the 2003 values for 20 of the 21 countries considered so far in an easy-to-read tabular form (see Table 5).

[13] The construction of new roads satisfying strict safety norms is however not captured by this indicator.

Table 5: Density of motorways (km/100km^2) (Eurostat, 2007)

	1995	2003
EU-25	1.2	
EU-15	1.4	
Belgium	5.5	5.7
Czech Rep	0.5	0.7
Denmark	1.8	2.4
Germany	3.1	3.4
Estonia	0.1	0.2
Greece	0.3	
Spain	1.4	2.0
France	1.5	1.9
Ireland	0.1	0.3
Italy	2.1	2.2
Cyprus	1.8	2.9
Latvia	-	-
Lithuania	0.6	0.6
Luxembourg	4.4	5.7
Hungary	0.4	0.6
Malta	-	-
Netherlands	5.3	6.1
Austria	1.9	2.0
Poland	0.1	0.1
Portugal	0.7	2.2
Slovenia	1.4	2.3
Slovakia	0.4	0.6
Finland	0.1	0.2
Sweden	0.3	0.4
Unt Kingdom	1.4	1.5
Bulgaria	0.3	0.3
Croatia	0.5	1.0
Romania	0.0	0.0
Turkey	0.2	0.2
Iceland	-	-
Liechtenstein	-	-
Norway	0.0	0.1
Switzerland	2.9	3.2

For Greece, another source is required. From the graphical representation of motorway density in 32 countries of the European Union Road Federation it is deduced that the value for Greece is similar to the one for Hungary. Consequently, value 0.6 (see Table 5) is assigned to Greece.

4.2.7 Trauma management

4.2.7.1 Relevant sources

For most of the trauma management indicators defined (e.g., percentage of calls to emergency medical services due to a road accident; average length of stay in the hospital after a road accident; share of road casualties who died during hospitalisation) data are collected by some bodies but no centrally maintained international data sources exist. In general, the following sources provide information on this final risk domain:

- The World Health Organization: total expenditure on health as share of the gross domestic product. Values can be obtained for a large set of countries with respect to different years. Moreover, other information (e.g., the number of physicians) can be found.
- Eurostat: the number of hospital beds per 10,000 citizens. 2003 values are available for 14 countries and 2002 values for 4 countries of the set of 21 countries.
- The SafetyNet project: the number of emergency medical services' staff per 10,000 citizens. Data are available for 15 countries (of which 3 not in the set of 21 countries focused on so far). Moreover, information has been collected on the number of EMS stations per 10,000 citizens and per 100 km of rural road length; the percentage of physicians and paramedics out of the total EMS staff; the number of EMS transportation units per 10,000 citizens, etc.

Several indicators from the World Health Organization and Eurostat can be considered. However, they are not specifically directed towards road safety, but more general indices of the health system in a country. Within the SafetyNet project, indicators related to emergency medical services are being defined. Although they have a clearer link with road safety, data are only available for a limited set of countries. As this would largely restrict the set of countries for which a road safety performance index score can be obtained the expenditure on health as share of the gross domestic product is the selected best available indicator for the trauma management domain. The data are presented next.

4.2.7.2 World Health Organization

Trauma management indicator values are obtained from the world health report of the World Health Organization. The 2003 values for countries in the European region are shown in Table 6.

Table 6: Total health expenditure as gross domestic product share (World Health Organization, 2006)

Country	2003	Country	2003	Country	2003
Albania	6.5	Germany	11.1	Portugal	9.6
Andorra	7.1	Greece	9.9	Rep of Moldova	7.2
Armenia	6.0	Hungary	8.4	Romania	6.1
Austria	7.5	Iceland	10.5	Russian Fed	5.6
Azerbaijan	3.6	Ireland	7.3	San Marino	7.5
Belarus	5.5	Israel	8.9	Serbia Montenegro	9.6
Belgium	9.4	Italy	8.4	Slovakia	5.9
Bosnia Herzegovina	9.5	Kazakhstan	3.5	Slovenia	8.8
Bulgaria	7.5	Kyrgyzstan	5.3	Spain	7.7
Croatia	7.8	Latvia	6.4	Sweden	9.4
Cyprus	6.4	Lithuania	6.6	Switzerland	11.5
Czech Rep	7.5	Luxembourg	6.8	Tajikistan	4.4
Denmark	9.0	Malta	9.3	The former Yugoslav Rep of Macedonia	7.1
Estonia	5.3	Monaco	9.7	Turkey	7.6
Finland	7.4	Netherlands	9.8	Turkmenistan	3.9
France	10.1	Norway	10.3	Ukraine	5.7
Georgia	4.0	Poland	6.5	Unt Kingdom	8.0
				Uzbekistan	5.5

4.3 Data sources on final outcomes

This research deals with the methodology for constructing a composite road safety performance index. The various methodological aspects are illustrated using safety performance data. In addition, data related to the final outcome level (such as the number of persons killed; see Section 1.5.2) are used in this research.

We can dispose of a number of sources in which final outcome data for a large set of countries are given. In the proposal of EuroRIS, i.e., the European Road safety Information System (SafetyNet, 2005b) information on CARE, EuroStat, ECMT, IRTAD, IRF, UNECE and WHO is provided. The Community database on Accidents on the Roads in Europe (CARE) contains disaggregated information on fatalities and fatal accidents over time for a large number of European countries. EuroStat annually publishes the EuroStat Yearbook – with data on population, health (e.g., fatalities from transport accidents), economy, etc in annual time series of 10 years – and EU Energy and Transport in Figures[14] containing information on EU member states and other countries (e.g., Switzerland and Japan). The European Conference of Ministers of Transport (ECMT), recently transformed into the International Transport Forum (ITF), publishes the annual Statistical Report on Road Accidents in 26 European countries. Data on the number of fatalities, casualties, injury accidents and motor vehicles are presented and the main road safety actions and changes in regulation are listed. The International Road Traffic and Accident Database (IRTAD) maintained by the Organisation for Economic Co-operation and Development contains information for 28 countries with respect to fatalities, seriously injured persons, population, network length, vehicle fleet, etc. Recently, IRTAD and ECMT (now ITF) have joined their forces by forming a joint OECD/ITF transport research centre. The International Road Federation (IRF) publishes at the end of each year an overview with data on several aspects of traffic and transport for more than 100 countries. The European Union Road Federation (ERF) publishes information on 27 European countries. In addition, the Economic Commission for Europe of the United Nations (UNECE) annually publishes a multitude of data among which traffic

[14] This is in cooperation with the Directorate-General for Energy and Transport of the European Commission.

safety related data. Finally, the World Health Organization (WHO) set up a mortality database with annual world health statistics on causes of death.

These data sources contain overlapping information but are not entirely consistent. For example, a value that is missing for a particular year in one database or publication is present in another one. However, this might not be an official value. Furthermore, elements present in more than one publication do not necessarily have the same value nor do they have the same value as published by the official national institute. Therefore, it is essential to know where the data come from and which procedures lie at the basis of their collection and publication (SafetyNet, 2005b).

Moreover, international projects set up around a specific road safety related theme often have a data collection component. Relevant road safety information has been collected in SafetyNet, SUNflower (+6), SARTRE, PENDANT, IMMORTAL, ALCOLOCK, CRASH TEST DATABASE, etc[15]. There exist a number of European databases containing data related to specific road user groups or locations. Examples are CHILD (children), ECBOS (coach and bus occupants), ETAC (truck accidents), MAIDS (motorcyclists) and RISER (highway accidents). Finally, the European Transport Safety Council publishes four times a year a ranking of countries in terms of their performance in a certain road safety area, e.g., motorways or motorcyclists.

Nowadays, comparisons of countries in terms of road safety often occur based on final outcome indicators. As shown by the previous paragraphs, final outcome data can be found in several sources. Most data relate to the number of fatalities rather than the number of serious casualties or the number of slight casualties. A larger amount of uncertainty is linked to these latter two variables as their definition often differs across countries. Moreover, the degree of underreporting is considered to be the lowest among fatalities (see e.g., Elvik and Vaa, 2004). Instead of using absolute numbers the benchmarking of countries differing with respect to size, population, vehicle fleet, etc requires the use of relative variables. Common examples are the mortality rate (per

[15] More information on these projects can be found via http://ec.europa.eu/transport/roadsafety/publications/projectfiles/alphabetically_en.htm.

population), the fatality rate (per vehicles) and the fatality risk (per vehicle kilometres).

Focusing on the three most common exposure measures – population, vehicle fleet and vehicle kilometres – the latter one can be expected to best represent the level of exposure as two countries with a same number of inhabitants or even a same number of registered vehicles might exhibit very different mobility patterns. Nevertheless, countries may use their own method to determine the number of vehicle kilometres, thereby causing extra uncertainty and less comparability. The use of population figures which are regularly and uniformly collected, is most popular. Exposure data with respect to 2003 can be deduced for a large set of countries from the IRTAD. Considering the set of 21 European countries, vehicle kilometre values are available for 10 countries whereas population values for 19 countries.

The number of road fatalities per million inhabitants will be used for the final outcome level because it is characterized by a high degree of comparability and data availability. The 2003 values presented in the report of the European Union Road Federation (2006) are used. Data are available for 20 of the 21 countries considered in this study, i.e., Austria (AT), Belgium (BE), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IE), Italy (IT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SL), Spain (ES), Sweden (SE) and United Kingdom (UK). For Switzerland (CH), rarely incorporated in European overviews, the number of fatalities per million inhabitants is deduced from the Organisation for Economic Co-operation and Development (2008). The final outcome data are visualised in Figure 14.

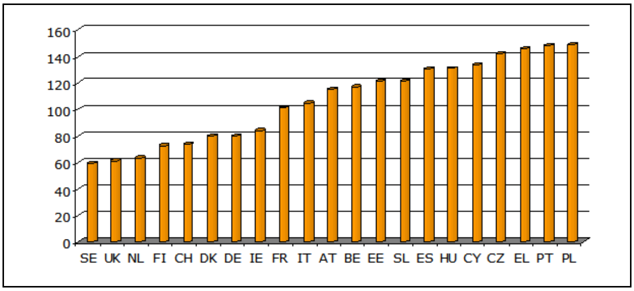


Figure 14: Road fatalities per million inhabitants in 21 European countries (2003)

4.4 Concluding remarks

Various international data sources providing indicator values for a large set of countries were consulted. While the search for additional and better data is an ongoing process, the index methodology will be illustrated using six indicators which we believe reflect essential and best available information to track road safety performance. Since no data are available for all indicators over some period of time, the composite road safety performance index will be computed for one year only. The focus is on the most recent year for which indicator data for 21 European countries are available at the time of writing, which is 2003.

At this point, some remarks regarding the finally selected indicator data set are given. First of all, the indicator listing and data collection started from seven risk domains considered in the most recent source with respect to road safety performance indicators (Vis, 2005). In addition to alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management other risk factors having a strong relationship with road safety, a large contribution to accidents and which can be influenced by measures (European Transport Safety Council, 2001) could be incorporated in the road safety performance index. Second, each risk domain is represented by one (best available) indicator. The index methodology is illustrated by combining domain indicators into a composite index (see Figure 12). A domain indicator can be a combination of specific indicators itself. Given the difficulties encountered in the identification of good road safety performance indicators and in order to avoid unnecessary complexities only the most appropriate or best available indicator is considered per risk domain at this moment. Third, proxy indicators and data are used. The maturity of safety performance indicators as well as the availability and quality of the corresponding data varies between the seven risk domains. The best available daytime running lights indicator already scored below average at the evaluation (in Appendix II) and appeared to be problematic. Therefore, no indicator for the daytime running lights domain is considered in the index construction process at this moment. Furthermore, the best available roads and trauma management indicator are proxy indicators for measuring the performance regarding these two risk aspects. Finally, we will make use of self-reported data with respect to alcohol and drugs as well as speed. It should be noted that self-reporting surveys might produce considerable response bias (e.g., Parada et al., 2001). In particular, in case the tendency of respondents towards socially acceptable answers differs between countries (due to their culture) self-reported values imply some kind of distortion of the differences prevalent in reality.

The main focus of this manuscript is to develop a scientifically sound and appropriate index methodology. Theoretical concepts will be illustrated for a large set of countries using a best available yet imperfect indicator data set. The indicator for the alcohol and drugs domain is the percentage of car drivers reporting to often drive while having a blood alcohol content above the legal limit, with data obtained from the Social Attitudes to Road Traffic Risk in Europe research. The percentage of car drivers reporting to often exceed the maximum speed limit in built-up areas is the chosen speed indicator. These data were also derived from SARTRE. As third indicator we select the percentage of persons wearing a seat belt in the front seats of a car or van with data from the European Transport Safety Council. Values for the share of relatively new cars (i.e., less than 6 years old), which is the indicator for the vehicle domain, were obtained from the Eurostat database. For the roads domain, the density of motorways is the best available indicator. Data were deduced from the Eurostat yearbook. Finally, the trauma management domain is represented by the share of the gross domestic product spent on health care using data from the World Health Organization.

Beside safety performance indicators final outcome indicators can be considered. In this research, the number of road fatalities per million inhabitants will be used. This common effect indicator considering on the one hand fatalities – the type of outcome most consistently defined and most completely registered – and on the other hand population figures – for which comparable, reliable data are available – is a good choice. The road safety performance index should be constructed in such a way that it has a clear link with the final outcome indicator. That way, the performance index and its indicators can act as warning signals for urgent action.

In the next chapter, the safety performance indicator data set is studied in more detail by means of univariate and multivariate analyses. However, before doing any analysis the direction of the indicators in terms of their expected road safety impact should be assessed and geared to one another. In other words, a high (or low) performance indicator value should always correspond to a low number of fatalities per million inhabitants. Here, we opt for a negative relationship in which high performance indicator values should be aimed at. Therefore, the direction of the alcohol and drugs indicator and the speed indicator should be adapted. The percentage of car drivers reporting to often drive while having a blood alcohol content *below* the legal limit respectively to often drive in built-up areas at a speed *below* the legal limit become the new indicators. As both indicators were expressed in terms of percentages the new values are obtained by simply subtracting the original SARTRE values from 100. That way, a country having the highest score on all safety performance indicators – the highest percentage of drivers respecting the legal alcohol limit, the highest percentage of drivers respecting the legal speed limit, the highest seat belt wearing rate, the highest share of 'young' cars, the highest motorway density and the highest GDP share for health expenditure – is considered to have the lowest number of road fatalities per million inhabitants (in the near future).

Chapter 5 Univariate and multivariate analyses

5.1 Introduction

The process of building a road safety performance index will be illustrated using the six best available indicators described in Chapter 4. More specifically, the following indicators will be combined: the percentage of car drivers reporting to respect the legal alcohol limit; the percentage of car drivers reporting to respect the maximum speed limit in built-up areas; the percentage of persons wearing a seat belt in the front seats of a car or van; the percentage of cars less than six years old; the density of motorways; and the expenditure on health as share of the gross domestic product.

In this chapter, the data of the six selected safety performance indicators are studied in more detail. First, each indicator is discussed individually by means of univariate analyses. Second, multivariate analyses studying the interrelationships between the different indicators are performed. The aim of these analyses is to gain insight into the structure of the indicator set and to assess whether appropriate indicators have been selected for combination in an index. Section 5.2 provides summary statistics on each indicator and visually presents the data, discusses common data transformation (normalisation) techniques and briefly describes how missing values can be dealt with. In Section 5.3, a number of analyses are performed in order to get an idea about the degree of association between the indicators, about the internal consistency of the indicators, about which indicators can be grouped together, about (dis)similar performance between countries and about the explanatory power of the selected indicators in terms of the number of road fatalities per million inhabitants. Finally, concluding remarks with respect to these descriptive analyses are given in Section 5.4.

5.2 One-dimensional exploration

5.2.1 Describing and visualising

The 2003 data for each of the six indicators for the 21 European countries considered in this research are described by means of some summary statistics (see Table 7) and graphically presented in Figure 15. In the table, values with respect to the central tendency and distribution are given. Focusing on the protective systems indicator it can be seen that on average in the 21 countries there is a seat belt wearing rate in the front seats of cars and vans of 79.6%. Moreover, the most frequent value in the data set is 75% while the median equals 82%. The intervals [40;75[, [75;82[, [82;88[and [88;97] contain one quarter of the observations. A variance of 175.8 or a standard deviation of 13.3 implies a considerable variability in the protective systems data. Finally, the skewness is a measure of the asymmetry of the distribution. Large negative values (i.e., >-1 in this case) imply that the distribution of the indicator has a long left tail whereas large positive values express a long right tail. In Figure 15, the data for each indicator are visualised.

Table 7: Summary statistics on the indicator data

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
# of cases	21	21	21	21	21	21
Mean	96.10	92.38	79.57	34.48	1.95	8.52
Modus	99.7	94	75	/	2	7.5
Median	97.5	93	82	36.74	2	8.4
25th percentile	94.9	89	75	28.69	0.6	7.5
75th percentile	98.7	94	88	39.43	2.4	9.6
Minimum	78.2	88	40	17.01	0.1	5.3
Maximum	100	97	97	53.67	6.1	11.5
Range	21.8	9	57	36.66	6	6.2
Variance	22.80	10.05	175.76	99.25	2.82	2.47
Stand deviation	4.78	3.17	13.26	9.96	1.68	1.57
Skewness	-2.84	-0.38	-1.42	-0.23	1.15	-0.04

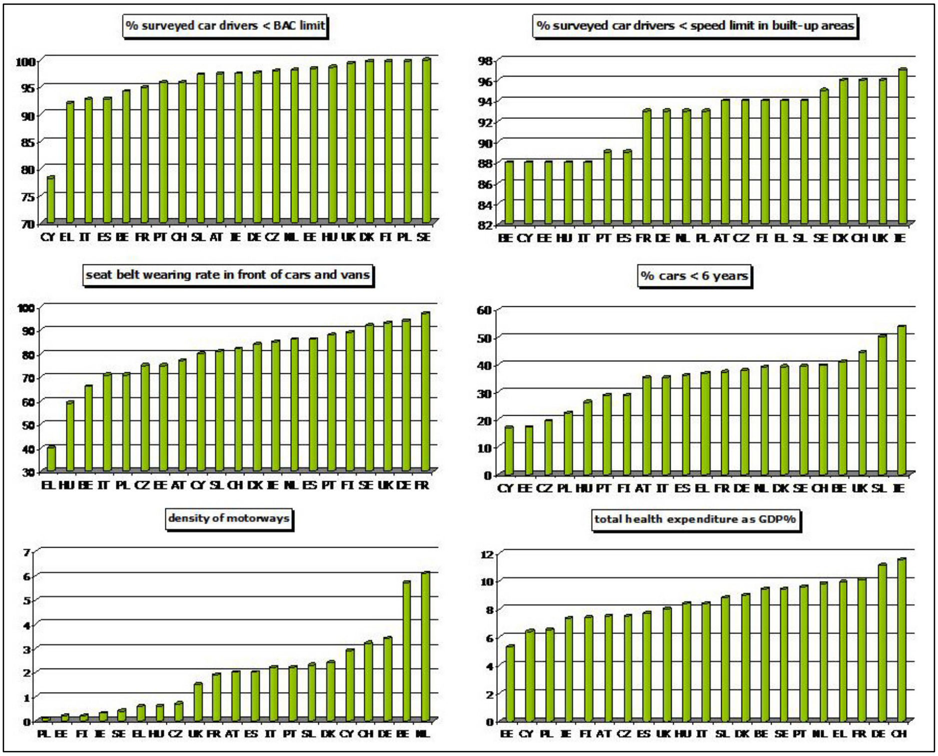


Figure 15: Values of the six selected indicators

Figure 15 shows that in 20 of the 21 countries (Cyprus being the exception) at least 90% of the respondents state to (at least often) respect the legal alcohol limit. For Sweden, the maximum score of 100% has been reported. Secondly, Ireland has the best speed performance. Thirdly, quite dissimilar seat belt wearing rates have been reported for the 21 European countries with a value of 40 for Greece and 97 for France. Next, the share of cars of maximum five years old lies between 17 (Cyprus) and 54 (Ireland). The Netherlands and Belgium are the two outperforming countries with respect to motorway density whereas this density is smaller than 1% in eight countries. Finally, the gross domestic product shares spent on health vary between 5 (Estonia) and 11.5 (Switzerland).

This section briefly discussed the raw indicator values of the 21 countries. Some indicators show a large variability and even extreme values. However, as all data have been deduced from reliable international sources (see Chapter 4) they are considered to be correct values that are suitable to be used rather than outliers needing some adjustment.

5.2.2 Normalising

Each indicator is expressed in its own measurement unit and presented on a different scale (see Figure 15). For the creation of a composite index it is essential to eliminate differences in size in order to avoid that one indicator dominates the final index score due to its relatively large values. Moreover, normalising the raw values results in easily understandable and comparable data. Various normalisation techniques exist (see e.g., Nardo et al., 2005b) of which three common and valuable methods will be used in this study.

Table 8 shows the rescaled indicator values. The difference between each raw value and the indicator minimum has been divided by the range of the indicator. Consequently, the worst or lowest indicator value is transformed into zero while the best or highest indicator value obtains a rescaled score of one. All rescaled values lie within this interpretable interval. Furthermore, countries with a good safety performance or in other words with relatively high rescaled values (i.e., close to one) can be detected at a glance. Finally, this normalisation technique transforms all raw values in such a way that the new range becomes one. Table 8 shows that Cyprus is the worst performing country with respect to three of the six indicators. Germany, the Netherlands and Switzerland are countries with rescaled values above 0.5 on all six dimensions.

Table 8: Rescaled data set

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
Austria	0.88	0.67	0.65	0.49	0.32	0.35
Belgium	0.73	0.00	0.46	0.65	0.93	0.66
Cyprus	0.00	0.00	0.70	0.00	0.47	0.18
Czech Rep	0.91	0.67	0.61	0.07	0.10	0.35
Denmark	0.99	0.89	0.77	0.60	0.38	0.60
Estonia	0.93	0.00	0.61	0.00	0.02	0.00
Finland	0.99	0.67	0.86	0.32	0.02	0.34
France	0.77	0.56	1.00	0.55	0.30	0.77
Germany	0.89	0.56	0.95	0.57	0.55	0.94
Greece	0.64	0.67	0.00	0.54	0.08	0.74
Hungary	0.94	0.00	0.33	0.25	0.08	0.50
Ireland	0.89	1.00	0.79	1.00	0.03	0.32
Italy	0.67	0.00	0.54	0.50	0.35	0.50
Netherlands	0.91	0.56	0.81	0.60	1.00	0.73
Poland	0.99	0.56	0.54	0.14	0.00	0.19
Portugal	0.81	0.11	0.84	0.32	0.35	0.69
Slovenia	0.88	0.67	0.72	0.90	0.37	0.56
Spain	0.67	0.11	0.81	0.52	0.32	0.39
Sweden	1.00	0.78	0.91	0.61	0.05	0.66
Switzerland	0.81	0.89	0.74	0.62	0.52	1.00
Unt Kingdom	0.97	0.89	0.93	0.75	0.23	0.44

Furthermore, the rescaled values are suitable for visualising all six performances of each country on a spider diagram. Each of the six axes of this diagram ranges from zero (the centre) to one (the edge). The longer the distance from the centre the better as a higher indicator value implies a better road safety performance. Rather than showing all countries' values in one graph the 21 countries have been divided into more or less equally sized groups based on their geographical position (see Figure 16). Denmark, Estonia, Finland, Ireland, Sweden and United Kingdom belong to the group of Northern countries (N); Belgium, France, Germany, the Netherlands and Switzerland are classified as Western countries (W); Austria, Czech Republic, Hungary, Poland and Slovenia are the Eastern countries (E); and the group of Southern countries (S) consists of Cyprus, Greece, Italy, Portugal and Spain.

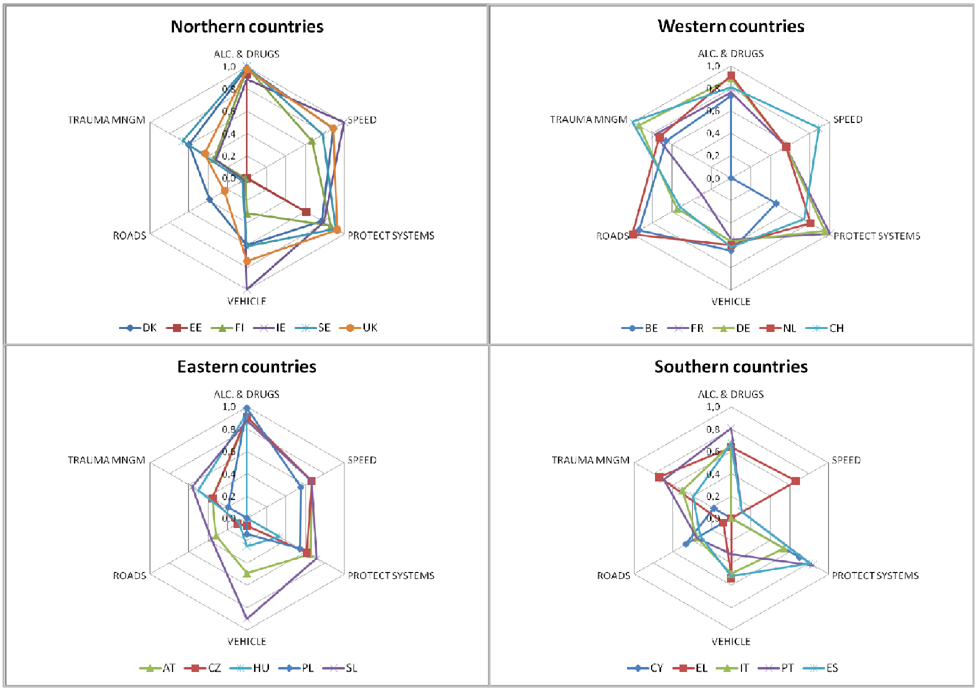


Figure 16: Spider diagrams of road safety performance per group of countries

On the one hand, Figure 16 shows some differences in performance on the six indicators between the four groups of countries; on the other hand differences between the countries within each group can be noted as well. On average, the Northern countries perform best with respect to the alcohol and drugs, speed and protective systems indicator and worst with respect to the roads indicator. The group of Western countries performs relatively good on all six aspects and best on the trauma management, roads and vehicle dimension. By contrast, the groups of Eastern and Southern countries are characterized by smaller scores. Of the four groups, the Eastern countries' group performs worst regarding the vehicle and protective systems indicator whereas the group of Southern countries has on average the lowest alcohol and drugs and speed value. In addition, the lines connecting the six rescaled indicator values of each country indicate some dissimilarities within each group. The best performing country of each group is Denmark (N), the Netherlands (W), Slovenia (E) and Portugal (S). The worst performing Northern, Western, Eastern and Southern country is Estonia respectively Belgium, Hungary and Cyprus.

Standardisation is a second important normalisation technique. A standardised value is equal to the difference between the raw value and the mean value divided by the standard deviation. Rather than indicators with a similar range standardisation transforms the raw values such that the mean of each indicator becomes zero and its standard deviation one. That way, extreme values can be easily detected as their coefficients are relatively large (e.g., -3.75). The 6×21 standardised values are presented in Table 9. Countries performing below average are characterized by a negative standardised value. Consequently, negative index scores are possible in case standardised values are combined.

Table 9: Standardised data set

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
Austria	0.27	0.51	-0.19	0.07	0.03	-0.65
Belgium	-0.40	-1.38	-1.02	0.64	2.23	0.56
Cyprus	-3.75	-1.38	0.03	-1.75	0.57	-1.35
Czech Rep	0.40	0.51	-0.34	-1.51	-0.74	-0.65
Denmark	0.75	1.14	0.33	0.47	0.27	0.30
Estonia	0.48	-1.38	-0.34	-1.74	-1.04	-2.05
Finland	0.75	0.51	0.71	-0.57	-1.04	-0.72
France	-0.25	0.20	1.31	0.27	-0.03	1.00
Germany	0.32	0.20	1.09	0.34	0.86	1.64
Greece	-0.84	0.51	-2.98	0.23	-0.80	0.88
Hungary	0.55	-1.38	-1.55	-0.82	-0.80	-0.08
Ireland	0.29	1.46	0.41	1.93	-0.98	-0.78
Italy	-0.71	-1.38	-0.65	0.08	0.15	-0.08
Netherlands	0.42	0.20	0.48	0.45	2.47	0.81
Poland	0.75	0.20	-0.65	-1.22	-1.10	-1.29
Portugal	-0.06	-1.07	0.64	-0.58	0.15	0.69
Slovenia	0.25	0.51	0.11	1.57	0.21	0.18
Spain	-0.69	-1.07	0.48	0.16	0.03	-0.52
Sweden	0.82	0.83	0.94	0.50	-0.92	0.56
Switzerland	-0.06	1.14	0.18	0.52	0.75	1.89
Unt Kingdom	0.69	1.14	1.01	1.00	-0.27	-0.33

The third normalisation method uses rank numbers. Instead of focusing on the exact indicator values, ordinal information is used. More specifically, the best performing country is assigned value 1 whereas value 21 is given to the country performing worst with respect to a specific indicator. In case countries have the same raw value, the mean rank is assigned (e.g., 15.5 instead of 15 and 16). The normalised indicator values based on rank numbers are presented in Table 10.

Table 10: Ordinal data set

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
Austria	12	8	14	14	10.5	15.5
Belgium	17	19	19	4	2	7.5
Cyprus	21	19	13	21	5	20
Czech Rep	9	8	15.5	19	14	15.5
Denmark	3	3	10	7	6	9
Estonia	7	19	15.5	20	19.5	21
Finland	3	8	5	15	19.5	17
France	16	12.5	1	10	12	3
Germany	10	12.5	2	9	3	2
Greece	20	8	21	11	15.5	4
Hungary	6	19	20	17	15.5	11.5
Ireland	11	1	9	1	18	18
Italy	19	19	17.5	13	8.5	11.5
Netherlands	8	12.5	7.5	8	1	5
Poland	3	12.5	17.5	18	21	19
Portugal	14.5	15.5	6	16	8.5	6
Slovenia	13	8	12	2	7	10
Spain	18	15.5	7.5	12	10.5	14
Sweden	1	5	4	6	17	7.5
Switzerland	14.5	3	11	5	4	1
Unt Kingdom	5	3	3	3	13	13

With this type of normalisation, on the one hand some information is lost as the properties of an equal-interval scale are no longer valid (e.g., the difference in the alcohol and drugs value between ranks 20 and 21 is extremely large while this difference between ranks 19 and 20 is very limited). On the other hand, attention is paid to the ordering of the countries which is expected to be relatively stable and less influenced by measurement error. At a quick glance the leaders and laggards can be identified. Germany, Sweden, Switzerland and United Kingdom have a top-five position on at least three of the six dimensions. In addition, Denmark, the Netherlands and Slovenia perform relatively well on all dimensions as their worst rank is 10 respectively 12.5 and 13. Cyprus, Estonia and Poland are characterized by very bad rank numbers on four dimensions.

To conclude, the aim of normalisation is to transform the original values of a set of indicators in order to make them comparable. Each method normalises in its own way thereby resulting in other normalised scores and even other end results (e.g., index scores). The three methods presented here – rescaling, standardisation and rank numbers – all have some interesting advantage. These normalised data sets will be used to create the road safety performance index. The impact of the chosen normalisation technique on the end result will be assessed in Chapter 9. The multivariate analyses in Section 5.3 will be performed on the normalised indicator values presented in Table 9 because the ratio of two numbers is best kept by this method. The final aspect with respect to univariate analysis – how to deal with missing values – is discussed in the next section.

5.2.3 Missing values

In Chapter 4, data have been collected for the best available indicators. Starting from the set of 23 European countries participating in the SARTRE research and therefore having both an alcohol and drugs and speed value, values for the other indicators were searched for the same set of countries. However, in the other data sources values for other sets of countries were found and for Croatia and Slovakia a large share of essential indicator information appeared to be missing. The issue of missing values is outside the scope of the current research and the creation of a road safety performance index will be illustrated for 21 countries. Nevertheless, as the development of road safety performance indicators as well as data collection efforts will increase in the future, it seems appropriate to briefly touch this subject as it will become a frequent problem. In case only a few indicator values are missing for a country some techniques can be applied enabling the researcher to consider this country in the analysis as well. When only countries with a(n almost) perfect road safety performance data system would be incorporated in the analysis the results would be misleading as underperforming countries are often characterized by a bad data system. Although the issue of missing values can be considered to be a research domain in itself the main considerations are given here.

The missing values literature is extensive and in rapid development. In general, there are three main approaches for dealing with missing values, namely case deletion, single imputation and multiple imputation (Nardo et al., 2005b). However, the deletion of cases with missing values can generate biased results. The other two approaches consider missing values as part of the analysis and therefore try to impute these values. The most common applications of single imputation – in which case for each missing value one value is imputed – are mean, median or mode substitution, regression imputation and expectation-maximisation imputation. For multiple imputation, the Markov Chain Monte Carlo algorithm is very common. Multiple imputation consists of three steps: imputation, analysis and pooling. In the first step, missing entries of the incomplete data set are filled in *m* times (*m* usually between three and ten). Imputed values are drawn from a distribution and this step results in *m* complete data sets. In the second step, each of the *m* data sets is analysed. Thirdly, the *m* analysis results are integrated into one final result. Unless the degree of missing information is very large, in most situations there is little advantage in producing and analysing more than a few imputed data sets (in terms of efficiency of the estimate) (Schafer, 2006). The main advantage of multiple over single imputation is that uncertainty is accounted for by creating different versions of the missing data and observing the variability between imputed data sets.

Schafer and Graham (2002) gave an overview of the historical development. 'Until the 1970s missing values were handled primarily by editing (for example case deletion). In 1976, Rubin developed a framework of inference from incomplete data that remains in use today. The formulation of the expectation-maximisation algorithm by Dempster, Laird and Rubin in 1977 made it feasible to compute maximum likelihood estimates in many missing data problems. About the same time, Rubin introduced the idea of multiple imputation. The creation of multiple imputations was facilitated by computer technology and new methods for Bayesian simulation developed in the late 1980s (Schafer, 1997). Nowadays, multiple imputation is considered to be the most correct method and has become increasingly attractive for researchers in several sciences.'

Furthermore, it needs to be understood why data are missing. Missing data mechanisms are commonly described as falling into one of three categories: missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR) (Wayman, 2003). When data are MCAR missing values do not depend on the variable of interest or any other observed variable in the data set. Missing values are no different from non-missing cases in terms of the analysis being performed. In case of MAR, missing values do not depend on the variable of interest but they are conditional on some other variables in the data set. NMAR implies that missing values depend on the values themselves. Data are missing in an unmeasured fashion, termed non-ignorable. For almost all imputation methods, data are assumed to be at least missing at random (MAR or MCAR) (Schafer and Graham, 2002; Wayman, 2003).

Beside the amount of different techniques (of which the multiple imputation Markov Chain Monte Carlo algorithm is to be preferred), a wide range of software programs is nowadays available (van Buuren, 2005), e.g., NORM software by Schafer (1999). The imputation of missing values is an interesting topic for further research.

5.3 Multidimensional exploration

When creating an index the interrelationships between the indicators to combine should be understood. Therefore, multivariate analyses are performed. Moreover, the results of these analyses provide insight into whether the indicators are appropriate for combining into an index. Using correlation analysis, pairs of indicators with a strong degree of association can be assessed (Section 5.3.1); Cronbach's alpha determines the internal consistency of the current data set (Section 5.3.2); principal components analysis groups the various indicators into a limited number of components (Section 5.3.3); similarly performing countries can be identified by means of cluster analysis (Section 5.3.4); and regression analysis is performed to link the indicators to the final outcome indicator (Section 5.3.5). Each technique is discussed below.

5.3.1 Correlation analysis

The first multivariate analysis technique deals with the correlation between each pair of indicators. For this research, non-parametric techniques not implying assumptions on the distribution of the data will be used. The degree of relationship between pairs of safety performance indicators as well as between each safety performance indicator and the final outcome indicator based on the Spearman's correlation coefficient is shown in Table 11.

Table 11: Spearman's correlation coefficients between six safety performance indicators and the final outcome indicator

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM	FATALITIES
ALC & DRUGS	1.000	0.390	0.251	0.037	-0.441	-0.204	-0.402
SPEED		1.000	0.363	0.588	-0.160	0.146	-0.495
PROTECT SYST			1.000	0.341	0.117	0.279	-0.615
VEHICLE				1.000	0.315	0.479	-0.609
ROADS					1.000	0.554	-0.187
TRAUMA MNGM						1.000	-0.293
FATALITIES							1.000

Four of the fifteen pairs of performance indicators appear to have a significant correlation at the 0.05 level: alcohol and roads; speed and vehicle; vehicle and trauma management; and roads and trauma management. The six dimensions used in this research are considered to each represent one particular aspect of road safety risk but at the same time some interrelationships between these dimensions can be expected.

The final column in Table 11 demonstrates the degree of correlation between each safety performance indicator and the number of road fatalities per million inhabitants. Recall that the performance indicators should have a clear link with road safety output in order to be valuable predictors based on which efficient policy measures can be taken. A country should aim at high performance levels and a low number of fatalities, therefore the six negative signs are as expected. Furthermore, the correlation is significant with respect to the speed, protective systems and vehicle indicator. The values for the roads and trauma management indicator are closer to zero which might be an indication of somewhat less appropriate performance indicators.

5.3.2 Internal consistency analysis

To assess the degree of internal consistency of a set of indicators Cronbach's alpha or the coefficient alpha can be used. With l = the number of indicators and x_0 = the sum of all indicators, it measures the share of total variability of the sample of indicators due to the correlation of the indicators (Nardo et al., 2005b):

$$\text{Cronbach's alpha} = \frac{l}{l-1}\left(1 - \frac{\sum_j \text{var}(x_j)}{\text{var}(x_0)}\right) \qquad j = 1,\ldots,l \qquad \text{(Eq. 5–1)}$$

Cronbach's alpha is most appropriately used when the indicators measure different substantive areas within a single construct. In case of no correlation between the indicators alpha equals zero. Usually, a value of 0.6 or less generally indicates unsatisfactory internal consistency (Malhotra and Birks, 1999). For the data set used in this research Cronbach's alpha is equal to 0.642. It can be stated that the selected indicators are more or less measuring the same underlying construct.

Furthermore, Cronbach's alpha can be computed if one indicator is deleted from the data set. This information is particularly useful in case the overall coefficient alpha results in a too small value. Then it can be deduced which indicator can best be eliminated in order to guarantee a higher level of internal consistency. From Table 12 it can be seen that a data set without the alcohol and drugs indicator results in a negligible increase in coefficient alpha. In case the roads indicator is no longer considered Cronbach's alpha would increase to 0.681. However, this gain in internal consistency is outweighed by the incorporation of (the best available indicator for) the roads domain in the safety performance index. For now, we conclude that the six selected indicators have a sufficient level of internal consistency.

Table 12: Cronbach's alpha for different sets of (six and five) safety performance indicators

	Cronbach's alpha
SIX INDICATORS	*0.642*
NO ALC & DRUGS	0.645
NO SPEED	0.567
NO PROTECT SYST	0.623
NO VEHICLE	0.496
NO ROADS	0.681
NO TRAUMA MNGM	0.549

5.3.3 Principal components analysis

The objectives of a principal components analysis are data reduction and interpretation (Johnson and Wichern, 2002). Indicators are grouped in a limited number of components still able to explain the majority of the variance in the data. The optimal number of principal components can be deduced from the software output (SPSS Inc., 2007). Several guidelines can be used in this respect (Sharma, 1996). Only components with an eigenvalue larger than one should be considered; only components which explain at least 10% of the total variance are to be chosen; the cumulative share of variance explained by the selected number of components should exceed 80%; etc. Apart from these rules, the scree plot – a visual representation of the eigenvalues and the component number – is an often used tool for assessing the optimal number of components. This number is taken to be the point at which the remaining eigenvalues are relatively small and all about the same size (Johnson and Wichern, 2002).

Based on the initial eigenvalues information in Table 13 and the scree plot presented in Figure 17 three principal components will be extracted. These three components account for almost 80% of the variance (the first component explains 38.1%, the second one 26.5% and the third one 14.7% of the total variance).

Table 13: Results of the principal components analysis

COMP	INITIAL EIGENVALUES			EXTRACTION SUMS OF SQUARED LOADINGS			ROTATION SUMS OF SQUARED LOADINGS		
	Total	*% of var.*	*Cumul.%*	*Total*	*% of var.*	*Cumul.%*	*Total*	*% of var.*	*Cumul.%*
1	2.288	38.141	38.141	2.288	38.141	38.141	1.870	31.174	31.174
2	1.589	26.482	64.622	1.589	26.482	64.622	1.852	30.860	62.034
3	0.879	14.656	79.278	0.879	14.656	79.278	1.035	17.244	79.278
4	0.573	9.547	88.825						
5	0.398	6.640	95.465						
6	0.272	4.535	100.000						

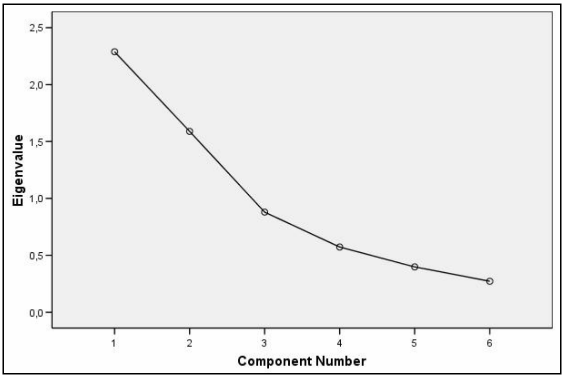


Figure 17: Scree plot

The initial component matrix (not shown here) consists of coefficients representing the correlations between the component and the indicators. Because the components are correlated with many indicators[16] this matrix seldom results in components that can be interpreted (Sharma, 1996). Therefore, through rotation, the component matrix is transformed into a simpler one that is easier to interpret. The very commonly applied varimax rotation procedure is used here since it minimizes the number of indicators with high coefficients on a component (Malhotra and Birks, 1999). The eigenvalues and the share of

[16] For the current data set, the first component is correlated with five of the six indicators (i.e., a coefficient higher than 0.5).

the variance explained by each component after rotation are given in the last columns of Table 13. Based on the coefficients in the rotated component matrix (not shown here) it can be determined that the alcohol and drugs indicator and speed indicator are grouped in component one; component two consists of the vehicle, roads and trauma management indicator; while the protective systems indicator represents component three. The grouping of the alcohol and drugs indicator with the speed indicator is rather reasonable whereas that of the vehicle, roads and trauma management component is more difficult to explain as they all represent a specific road safety risk aspect. In general, it can be stated that the use of a dimension reduction technique such as principal components analysis would be more suitable for a large set of indicators (e.g., several indicators for each risk domain). Nevertheless, this technique provided insight into the interrelationships present in the current data set.

5.3.4 Cluster analysis

The next multivariate analysis technique informs about the similarity in performance of the countries in the data set. A division in hierarchical clustering on the one hand and non-hierarchical clustering on the other can be made. As the latter requires a pre-specified number of clusters in which the countries should be classified hierarchical clustering is considered first. A dissimilarity matrix is set up using Ward's method (see e.g., Johnson and Wichern, 2002). At first, there are 21 clusters, each representing one country. In the first stage of the clustering process the two countries with the lowest coefficient (i.e., Czech Republic and Poland) are grouped. Next, the countries with the second lowest coefficient are grouped, and so on, until one cluster containing all 21 countries is obtained (at stage 20). The optimal solution within the two extremes (21 clusters and 1 cluster) needs to be found. Some guidelines are available for deciding on the number of clusters (see e.g., Malhotra and Birks, 1999). For hierarchical clustering, the distances at which clusters are combined can be used as criteria. This information can be obtained from the agglomeration schedule (table) or from the dendrogram (graph). The dendrogram is presented in Figure 18. In general, the relative sizes of the clusters (i.e., the number of countries belonging to each cluster) should be meaningful. The amount of clusters needs to be small but at the same time clusters need to represent similarly performing countries. Here, we opt for four clusters as they are easy to overlook. Cyprus appears to be performing fairly differently as it is not grouped with any other country. Moreover, in case more than four clusters are selected Greece would also be classified separately.

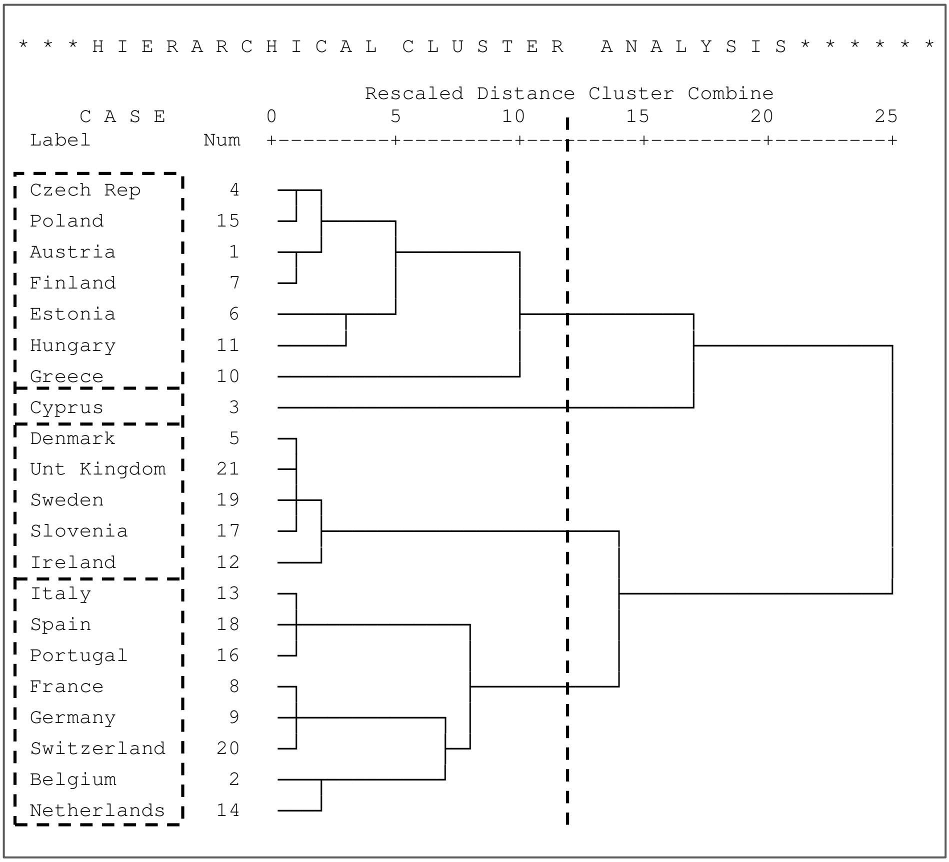


Figure 18: Dendrogram based on hierarchical clustering

A rescaled distance of twelve divided the 21 countries into four groups. This result from the hierarchical clustering will be used as input for the non-hierarchical clustering. Applying *k*-means clustering, the countries are classified in homogeneous groups that are characterized by a low within-variance and a high between-variance. The *k*-means clustering algorithm results in the following four classes:

- Austria – Denmark – Finland – France – Ireland – Slovenia – Sweden – United Kingdom (Cluster 1)

- Belgium – Germany – Greece – Italy – the Netherlands – Portugal – Spain – Switzerland (Cluster 2)
- Cyprus (Cluster 3)
- Czech Republic – Estonia – Hungary – Poland (Cluster 4)

The above classification again shows the very distinctive performance of Cyprus in terms of the six performance indicators. The strengths and weaknesses of each cluster can be deduced from the plotted cluster centres in Figure 19. Cyprus scores worst on four of the six risk domains. Cluster 1 has the highest cluster value concerning speed, protective systems and vehicle while cluster 2 (also consisting of eight countries) scores best on the roads and trauma management indicator. Finally, cluster 4 performs worst in terms of protective systems and roads. Of the six indicators vehicle, trauma management and speed appear to be particularly influential in determining the differences between the clusters.

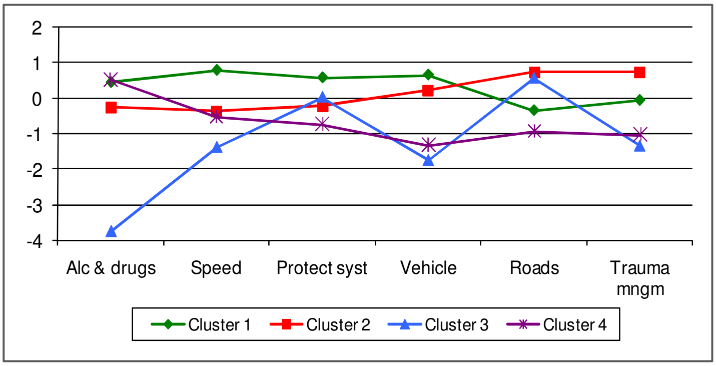


Figure 19: Cluster centres per country class

5.3.5 Regression analysis

Finally, the relationship between the six safety performance indicators on the one hand and the final road safety outcome indicator on the other is investigated. We are interested in the explanatory power of the selected performance indicators. Therefore, linear regression models are built with the number of road fatalities per million inhabitants as dependent variable *(Y)* and the six performance indicators as independent variables (X_1 = the alcohol and drugs indicator; X_2 = the speed indicator; X_3 = the protective systems indicator; X_4 = the vehicle indicator; X_5 = the roads indicator; X_6 = the trauma management indicator). First, all six indicators are entered in the regression model. Second, a stepwise selection method is applied in order to incorporate only significant indicators. Third, (first-order) interaction effects between the six performance indicators are taken into account; a stepwise selection method is also used here. In the construction of the regression models the satisfaction of the assumption of normality, no autocorrelation and homoscedasticity (Neter et al., 1996) has been checked by means of graphs. Moreover, from all possible models generated based on stepwise selection the presented models (see regression equations 2 and 3 below) are the ones with the highest adjusted R square resulting in significant variables and acceptable variance inflation factors (Neter et al., 1996). In Table 14, the results are given for each of the three regression models, i.e., the adjusted R square as well as the standardised coefficient, t-value and significance level of the independent variables incorporated in the model. The model fit and the significant variables offer valuable information. Moreover, the standardised coefficients can be compared to assess the most important variable(s). Since high safety performance indicator values and a small number of road fatalities per million inhabitants are to be aimed at negative coefficients are expected.

Entering all six indicators in the regression model results in an adjusted R square of 0.439. The six indicator coefficients are all negative. There is only one significant indicator, i.e., the protective systems indicator (p=0.04). Moreover, the trauma management indicator appears to be the least significant one. In the second regression model, a stepwise selection of the indicators occurred resulting in an increase in the adjusted R square (to 0.466). Both the protective systems (p=0.01) and the vehicle indicator (p=0.02) are incorporated in the model as significant explanatory factors for the number of road fatalities per million inhabitants.

In the third regression model significant interaction effects are present as well, resulting in a high adjusted R square (0.860). The final model includes four indicators – the speed, protective systems and roads indicator being significant – and three significant interaction effects, i.e., between the alcohol and drugs & the protective systems indicators, between the speed & the protective systems indicators and between the speed & the roads indicators. In other words, the change in the number of road fatalities per million inhabitants due to for example a change in speed also depends on the protective systems and roads values.

Table 14: Results of the regression analysis

	REGRESSION EQUATION AND OUTPUT		
1)	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$		
Adjusted R^2	0.439		
	Stand. Coeff.	*t value*	*Sig.*
(Constant)		2.905	0.012
Alc & drugs	-0.162	-0.821	0.425
Speed	-0.292	-1.219	0.243
Protect Syst	-0.404	-2.229	0.043
Vehicle	-0.181	-0.787	0.444
Roads	-0.212	-0.925	0.371
Trauma mngm	-0.010	-0.044	0.966
2)	$Y = \beta_0 + \beta_1 X_3 + \beta_2 X_4$		
Adjusted R^2	0.466		
	Stand. Coeff.	*t value*	*Sig.*
(Constant)		7.475	0.000
Protect Syst	-0.495	-2.944	0.009
Vehicle	-0.422	-2.511	0.022
3)	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_2 X_5$		
Adjusted R^2	0.860		
	Stand. Coeff.	*t value*	*Sig.*
(Constant)		39.027	0.000
Alc & drugs	0.038	0.338	0.741
Speed	-0.372	-3.151	0.008
Protect Syst	-0.570	-5.497	0.000
Roads	-0.235	-2.569	0.023
Alc & drugs X Protect Syst	-0.552	-4.876	0.000
Speed X Protect Syst	-0.597	-5.921	0.000
Speed X Roads	-0.257	-2.676	0.019

From the regression analysis it can be inferred that the six selected indicators already explain some part of the variance in the number of road fatalities per million inhabitants. As could be intuitively expected, the complexities involved in road safety cannot be entirely captured by six indicators, but requires a larger set of appropriate indicators. In general, it should be kept in mind that the number of observations is small compared to the number of variables. Moreover, the results presented here refer to the data used and should not be generalised. Possibly, another set of indicators, countries or year would yield different conclusions.

5.4 Concluding remarks

In this chapter, some first analyses have been performed on the data set consisting of six best available road safety performance indicator values for 21 European countries. In this step of the index process more insight into each indicator separately as well as into the structure and interrelationships of the entire indicator set is gained. Based on the analysis results it can be assessed whether the selected indicator set is appropriate for combination in an index. If this is not the case, other indicators should be selected (still scoring well on the selection criteria discussed in Section 3.2), data should be collected again and new analyses performed.

The indicator data have been presented graphically and described by means of summary statistics. To make the data comparable over the indicators, the normalisation of the original values has been handled. Attention was paid to three common methods, i.e., rescaling, standardisation and rank numbers. As they all have some advantage, these normalised values will be used in the remainder of this research. Subsequently, some (theoretical) considerations on the issue of missing values were discussed.

Additionally, several multivariate analysis techniques have been discussed. The correlation analysis resulted in four significant pairs of safety performance indicators. The internal consistency of the six indicators has been checked by means of Cronbach’s alpha. It appeared that the indicators are more or less measuring the same underlying construct. The principal components analysis grouped the six indicators in three components – alcohol and drugs & speed; vehicle & roads &

trauma management; protective systems – explaining 79% of the variance. By means of cluster analysis, the set of 21 European countries has been divided into four similarly performing classes. Cyprus appeared to have an extremely different performance on the six indicators and was not grouped with another country. Finally, by building some regression models it could be concluded that the six selected indicators already explain some part of the variance in the number of road fatalities per million inhabitants.

As the results from the multivariate analyses are satisfactory, i.e., the indicators have some level of correlation, internal consistency, clustering and explanatory power, the next step in the construction of a road safety performance index, i.e., the weighting of the indicators, will be discussed in Chapter 6.

Chapter 6 Assigning indicator weights[17]

In this chapter, we focus on a crucial step in the construction process of a composite road safety performance index, i.e., the assignment of weights to the individual indicators. Given the various weighting methods applied in literature (see e.g., Saisana and Tarantola, 2002) this chapter aims to provide insight into a number of common and valuable weighting methods for the road safety case. First, some essential theoretical considerations are given (Section 6.1) followed by the results of applying the methods on the (standardised) road safety indicator data set (Section 6.2). Finally, concluding remarks in terms of the most justifiable weighting method(s) for the construction of the road safety performance index are given in Section 6.3.

6.1 Weighting methods

In this section, the theory behind five weighting techniques is discussed. More specifically, weighting based on factor analysis (Section 6.1.1), analytic hierarchy process (Section 6.1.2), budget allocation (Section 6.1.3), data envelopment analysis (Section 6.1.4) and equal weighting (Section 6.1.5) will be handled.

6.1.1 Factor analysis

The first weighting method is based on factor analysis (FA). A factor analysis is often used to reduce the dimensions of a problem (Sharma, 1996). For the current road safety case, the six (I) dimensions are transformed into a smaller number of factors $(p < I)$ explaining the amount of correlation among the indicators. Having decided to consider p factors – based on the eigenvalue, the (cumulative) share of variance explained and/or the scree plot (see e.g., Sharma, 1996) – rotation is applied next. By applying varimax rotation (see also Section 5.3.3) a factor solution is obtained in which each factor is only correlated with a few indicators and in which each indicator (i) has a high loading on one factor (j) only (Sharma, 1996). Indicator weights can be deduced from

[17] Related research has been published in: Hermans, E., Van den Bossche, F. and Wets, G. (2008b). Combining road safety information in a performance index. Accident Analysis and Prevention, (40) pp. 1337-1344.

these rotated factor loadings by means of relatively limited computation. The simplified but equivalent approach to the calculations suggested in Nardo et al. (2005b) consists of the following steps:

- Define $u_{ij} = a_{ij}^2 / \sum_{m=1}^{l} \sum_{n=1}^{p} a_{mn}^2$ with rotated factor scores *a*.
- The preliminary weight of indicator i, $u_i = \max_j(u_{ij})$. However, the sum of the weights, $U = \sum u_i < 1$ due to the reduction of the dimensions.
- The final weight for each indicator i, $w_i = u_i / U$ and, by construction, $W = \sum w_i = 1$.

The use of factor analysis in the composite indicators field (e.g., the e-business readiness index in Pennoni et al., 2005) is not rare. However, this technique is often used to examine the interrelationships between the indicators instead of determining weights. The most important drawback is that weights are based on correlations which do not necessarily correspond to the real-world links between the phenomena being measured (Saisana and Tarantola, 2002). In addition, deducing weights from factor analysis requires a certain level of correlation in the data set (to reduce the problem in a number of factors), a justified selection of the optimal number of factors (as the weights depend on the chosen number of factors) and clear rotation results (because only the highest rotated factor loadings are used in the computation of weights). To conclude, this weighting method is most valuable in case several (sufficiently correlated) indicators per risk domain are considered.

6.1.2 Analytic hierarchy process

Analytic hierarchy process or AHP is a method developed by Saaty in the early 1970s in the field of decision theory in which a complex problem is translated into a hierarchy consisting of an overall goal, several (sub)criteria contributing to this goal and a number of alternatives to evaluate (Haas and Meixner, 2006). Both quantitative and qualitative criteria can be handled. With respect to deducing weights for the road safety performance indicators, experts are asked to judge the relative contribution of each indicator (criterion) to road safety compared to another indicator. They answer the questions 'which one of the two is more contributing to the overall goal (enhancing the level of road safety)?' and 'how large is the intensity of the difference?'. Values are given on a scale of 1 to 9. Equal contribution results in value 1, while 3 implies a slightly higher, 5 a strongly higher, 7 a very strongly higher and 9 an absolutely higher contribution of one indicator compared to another (Saaty, 1980). Having I indicators to judge, only $I(I-1)/2$ pairs have to be considered. The expert information can be presented in a reciprocal squared matrix with I rows and I columns consisting of values within the interval [1/9, 1/8, ..., 1, 2, ..., 9].

From the matrix containing expert information, the eigenvector with the largest eigenvalue (λ_{max}) has to be found. The eigenvector determines the weights and the eigenvalue is a measure for the consistency of the judgement. The consistency reflects the soundness of judgement or whether the interdependencies of the criteria are understood (Talbert et al., 1994). It is advisable to keep the number of criteria small and to define independent or at least sufficiently different criteria. Saaty (1980) defines the consistency index as $(\lambda_{max}-1)/(I-1)$. In addition to the consistency index, the consistency ratio is defined as the consistency index divided by the so-called average random index for a matrix of the same size[18]. A consistency ratio equal to zero corresponds to a perfect consistency while a value of one indicates meaningless (or random) estimates. As a rule-of-thumb, a consistency ratio smaller than 0.10 is considered to indicate satisfactory consistency. Such small ratios do not drastically affect the weights (Nardo et al., 2005b; Saaty, 1980). Besides inconsistency, subjectivity is a characteristic of the AHP method making the selection of the expert panel crucial. Judgement is affected by experience, depth of knowledge, relative intelligence, personal involvement, etc (Saaty, 1980).

To conclude, AHP is a comprehensible and popular technique that has already been used in the indoor environment index (Chiang and Lai, 2002) and the index of environmental friendliness (Puolamaa et al., 1996). Taking into account some level of subjectivity and inconsistency,

[18] The average random index value for a matrix is given in Saaty, 1980.

the information from well-selected experts is valuable for deducing indicator weights. In case of several experts, a number of possibilities exist to come to one final set of weights. This set of weights can reflect the opinion of one randomly selected expert or incorporate all (consistent) experts. In the latter case, the average or the median can be calculated or the group of experts could vote or reach consensus after a debate.

6.1.3 Budget allocation

Budget allocation (BA) is another method for obtaining indicator weights. A selected panel of experts is asked to distribute a given budget over the indicators in such a way that spending more on an indicator implies that they want to stress its importance. The weights can be obtained from a simple ratio.

In general, the BA method has four phases (Nardo et al., 2005b). First, the experts have to be selected. It is important to gather experts with a wide spectrum of knowledge and experience. Second, each expert allocates the predetermined budget of N points to the indicators. In a third step, weights are calculated from these figures. More specifically, the share of budget allocated to an indicator equals its weight. The fourth step is an optional one in which the procedure is iterated until convergence is reached.

Budget allocation is a simple and often used technique – see e.g., the e-business readiness index (Pennoni et al., 2005) and the internal market index (Tarantola et al., 2004) – with some limitations. Again, the selection of experts is crucial and should be well-considered. It is possible that the results are biased if an expert assigns a high weight to a dimension on which his/her country performs well. Moreover, the method may not measure the importance of a specific indicator but the need for political intervention in that dimension (Nardo et al., 2005b). Finally, the maximum number of indicators over which to distribute the budget is limited to ten, enabling the expert to keep an overview (Saisana and Tarantola, 2002). More information on the expert selection process and the resulting (budget allocation and analytic hierarchy process) weights for this research are given in Section 6.2.

6.1.4 Data envelopment analysis

Data envelopment analysis (DEA), developed by Charnes et al. (1978), is a performance measurement technique that can be used for evaluating the relative efficiency of decision making units (DMUs). For each DMU – country in our case – the efficiency is defined as the ratio of the weighted sum of outputs to the weighted sum of inputs (Cooper et al., 2000). Thereby, best weights are determined resulting in the most optimal objective value for a country. This implies that dimensions on which the country performs relatively well get a higher weight.

Translating the original DEA context to the composite index field implies that we do not consider inputs and refer to each indicator as an output. A general DEA model for indexes has been proposed in Cherchye et al. (2006); see (Eq. 6–1). For a particular country j in the data set $(j = 1,..,n)$ a composite index score is determined using a set of best indicator weights w_{ij} $(i = 1,..,I)$ which maximizes the composite index value of the country and satisfies the imposed restrictions. As stated by the second restriction in (Eq. 6–1) all weights are restricted to be non-negative. The first restriction guarantees an intuitive interpretation of the composite index and implies that no country in the data set can be assigned an index value larger than one under these weights.

$$CI_j = \max_{w_{ij}} \sum_{i=1}^{I} y_{ij} w_{ij} \qquad \text{(Eq. 6–1)}$$

$$\text{subject to} \quad \sum_{i=1}^{I} y_{ik} w_{ij} \leq 1 \qquad \forall k = 1,\ldots,n$$

$$w_{ij} \geq 0 \qquad \forall i = 1,\ldots,I$$

with CI = composite index score

j = country under study

w_{ij} = weight of indicator i of country j

I = number of indicators

y_{ij} = value of indicator i of country j

k = one of the countries in the data set

n = countries in the data set

For each country, a composite index score between zero and one can be obtained, with higher values indicating a better relative performance.

From the index scores, the best performing[19] countries – having an objective value of one – can be deduced. At the same time, underperforming countries will be revealed. For these countries with an optimal composite index score smaller than one, the country-specific weights can identify the problem areas. If the index score of a country is improved most in case of a better alcohol performance (i.e., that country has a low weight attached to alcohol), this information can be translated in specific alcohol action plans. Additionally, for underperforming countries the indicator values of the best performing countries can be used for target setting. Moreover, the different indicators can be listed in terms of priority. These aspects are handled in Chapter 7.

Compared to the previously discussed weighting methods, DEA is different. DEA is a method that can handle raw values making normalisation redundant. However, the weights do not sum up to one, which makes the comparison with weights from other methods impractical. Furthermore, a separate model is constructed for each country resulting in country-specific weights instead of one set of indicator weights for all countries. Therefore, we will next define one model for all countries in the data set in which the sum of the 21 composite index values is maximized, resulting in the same set of weights for all countries.

Another adaptation to the model presented in (Eq. 6–1) is the inclusion of extra restrictions on the weights. Total flexibility has been criticized on several grounds as the weights prove to be inconsistent with prior knowledge or accepted views (Pedraja-Chaparro et al., 1997). To obtain realistic weights (rather than e.g., all weight assigned to only one indicator) a multitude of ways to capture value judgements or prior information in a data envelopment analysis model is proposed in literature (Allen et al., 1997; Allen and Thanassoulis, 2004; etc). Cherchye et al. (2006) suggest to constrain the contribution of each indicator pie share (i.e., the product of the indicator value and the indicator weight) to the composite index value based on expert information obtained by means of budget allocation. The average of the two highest weights

[19] In the original DEA context, a distinction between 'efficient' and 'inefficient' countries is made. In this manuscript, we will use the terms 'best performing' and 'underperforming' countries.

assigned by the group of experts to each indicator (U_m) makes up the upper limit for the contribution of each pie share to the overall composite index score[20]. The DEA model that will be used for determining indicator weights is presented in (Eq. 6–2).

$$\sum_{j=1}^{n} CI_j = \max_{w_i} \sum_{i=1}^{l} \sum_{j=1}^{n} y_{ij} w_i \qquad \text{(Eq. 6–2)}$$

$$\text{s.t.} \quad \sum_{i=1}^{l} y_{ik} w_i \le 1 \qquad \forall k = 1,\ldots,n$$

$$\frac{y_{mj} w_m}{\sum_{i=1}^{l} y_{ij} w_i} \le U_m \qquad \forall m = 1,\ldots,l;\ \forall j = 1,\ldots,n$$

$$w_i \ge 0 \qquad \forall i = 1,\ldots,l$$

with CI = composite index score
j = country under study
n = countries in the data set
w_i = weight of indicator i irrespective of the country
l = number of indicators
y_{ij} = value of indicator i of country j
k = one of the countries in the data set
m = one of the indicators
U = upper limit

To conclude, DEA is a performance measurement technique in which the most favourable weights are selected both satisfying the imposed restrictions and resulting in the most optimal score. The results are influenced by the countries in the data set, hence this approach is only about relative performance (Anderson, 2006). Despite the original meaning this technique can be translated to be useful for composite indexes. This weighting method has already been used for a number of indexes (Cherchye et al., 2004; Cherchye et al., 2006) and to a limited extent in the road safety context (Odeck, 2005). Its strongest point is

[20] In the same way, lower limits could be defined. However, for the current indicator data set no feasible solution can be obtained in which the weights respect a combination of minimum and maximum limits for each indicator. More specifically, due to the extreme difference between the maximum and minimum value for the roads indicator (see Table 7) it is impossible to simultaneously meet the lower and upper limit for each country. Therefore, only upper limits are incorporated as they avoid the dominance of a particular indicator and take some level of relative importance into account.

that the weights are endogenously determined and derived directly from the data (Cooper et al., 2000). The best possible weights can be obtained for each country individually (Eq. 6–1) or for all countries in the data set (Eq. 6–2). Moreover, additional restrictions (e.g., based on expert information) can easily be incorporated leading to more acceptable weights. The presented DEA model is most valuable when some expert opinions are available and there is no agreement on the correct set of weights.

6.1.5 Equal weighting

In case of equal weighting, the same weight is assigned to each indicator. Since the weights sum up to one each indicator weight equals 1// (with / the number of indicators in the analysis). In case of several indicators per category (or risk domain) it is possible to use equal weighting for the main categories and for all indicators in the categories. That way, weights depend on the number of indicators in each category.

Equal weighting is a simple technique that is used in the creation of a number of composite indexes, e.g., the European innovation scoreboard (European Commission, 2004b) and the environmental sustainability index (Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network, 2005). Yet, it has some limitations. The most important drawback is that no insight is gained into the difference in importance of the indicators. As a result, equal weighting is not of great value for policymakers (nor researchers). In addition, it is unlikely that the resulting weights are similar to the real, unknown weights. Moreover, when two or more indicators are measuring the same, there is a risk of double weighting (Directorate for Science Technology and Industry, 2003). We conclude that equal weighting is a solution in case no other weighting method yields valid results. This approach works best if all indicators are uncorrelated or if they are all highly correlated.

6.1.6 Summary on the weighting methods

So far, five weighting techniques have been described. This section concludes with a summary of the characteristics of each weighting method. More specifically, Table 15 gives for each method the main advantages, the main disadvantages, the main requirements, the type of information used and an index in which the weighting method was applied.

Apart from the equal weighting method weights varying between the indicators are obtained. Data and/or expert opinions are used in this respect. This implies a more realistic composite index on the one hand and some level of subjectivity on the other. Regarding the data a best available yet imperfect data set is used: only six risk domains are considered, each risk domain is captured by one indicator and some proxy indicators and self-reported data are to be used at this moment (see also Section 2.9). Regarding the use of expert panels the related accuracy of the findings needs to be kept in mind (see e.g., Washington et al. 2009). More information on the selection of experts and the resulting weights is given in Section 6.2.1.

Table 15: Summary information on five weighting methods

	MAIN ADVANTAGES	MAIN DISADVANTAGES	MAIN REQUIREMENTS	INFORMATION	APPLICATION
Factor Analysis	- indicators are grouped in factors	- weights based on correlations may differ from reality	- some correlation between indicators - justification of the selected number of factors - clear rotation results	- data	- e-business readiness index (Pennoni et al., 2005)
Analytic Hierarchy Process	- detailed expert information - quantitative and qualitative criteria	- inconsistency - subjectivity	- carefully selected group of experts (with time) - small number of criteria - sufficiently different criteria	- experts	- index of environmental friendliness (Puolamaa et al., 1996)
Budget Allocation	- comprehensible - easy computation	- weight may indicate need for intervention - weight may represent dimensions a country performs on well	- carefully selected group of experts - maximum number of indicators is ten	- experts	- internal market index (Tarantola et al., 2004)
Data Envelopment Analysis	- best weights derived from data - value judgements can be included - no normalisation needed	- results are relative, i.e., influenced by the countries in the data set - sum of weights is not one - difficult to compare with other methods	- value judgements to obtain realistic weights - several countries in the data set	- data + experts	- technology achievement index (Cherchye et al., 2006)
Equal Weighting	- simple	- no insight in indicator importance - no added value for policymakers - risk of double weighting	- no valid results from other weighting methods - all indicators uncorrelated or highly correlated	- none	- European innovation scoreboard (European Commission, 2004b)

6.2 Results and discussion

The previous section described five common and useful methods for obtaining indicator weights. Each method will be applied to the standardised road safety indicator data set presented in Table 9 and evaluated by comparing the different results in terms of weights and rankings.

6.2.1 Indicator weights

A first set of indicator weights is obtained using factor analysis. The selected number of factors is three[21]. The alcohol and drugs indicator and the speed indicator represent factor 1; the vehicle, roads and trauma management indicator belong to factor 2; and factor 3 constitutes of the protective systems indicator. The (six times three) rotated factor loadings from the output are used to determine the indicator weights (as explained in Section 6.1.1). The indicator weights based on factor analysis are presented in the first row of Table 17 (on page 137). It can be seen that the protective systems indicator obtains the highest weight (0.232) and the vehicle indicator the lowest (0.110).

For the second and third method, the participation of experts was required. A panel of recognized experts with specific experience in the topic and a geographical representation was selected (Washington et al., 2009). More specifically, a group of experts participating in the European SafetyNet project was asked to cooperate. An international perspective on the contribution of each risk domain to road safety was obtained as the experts originated from different countries[22]. They gave their opinion independently from each other. Here, the average weights will be used in the computation of the index scores. However, Figure 20 and Figure 21 provide information on the variability in weights between the experts. The minimum, maximum and average weights as well as the standard deviations are shown. Furthermore, the impact of selecting the weights from one particular expert will be assessed in Chapter 9.

---

[21] In total, the three factors explain 80% of the variance; each of the three factors has a large contribution to this (>15%) and is characterized by an eigenvalue exceeding or close to one.

[22] i.e., Czech Republic, France, Greece, Hungary, Israel, the Netherlands, Norway, Switzerland and United Kingdom.

Regarding the analytic hierarchy process, each expert was asked to indicate the relative road safety contribution of each pair of risk domains. Results in terms of a matrix with [1/9, 1/8, ..., 1, 2, ..., 9] values were obtained from twelve experts. Using the Expert Choice software (Expert Choice Inc., 2006) consistency ratios and weights were determined based on each matrix. In order to obtain one set of appropriate AHP weights, the weights resulting from the nine experts with an acceptable degree of consistency (i.e., a consistency ratio of maximum 0.1) are averaged. As shown in Figure 20 and the second row of Table 17 most weight is assigned to the speed domain and the alcohol and drugs domain. Moreover, the graph shows that the nine experts disagree with respect to protective systems and roads whereas they have more or less the same opinion about the contribution of the trauma management domain.

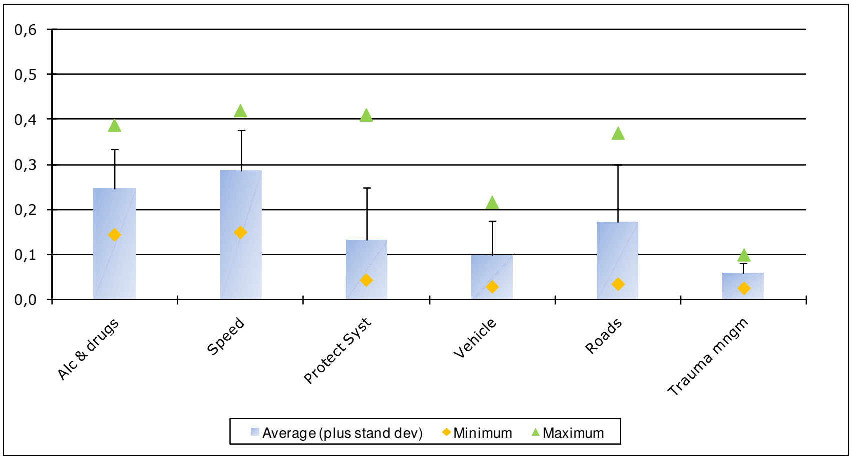


Figure 20: Analytic hierarchy process weights from a panel of experts

With respect to the budget allocation method, valid results were obtained from eleven experts. Based on the share that each expert assigned to the six domains indicator weights could be computed. The minimum, maximum and average weights as well as the standard deviation are shown in Figure 21. The experts agree most with respect to the vehicle domain to which the least budget has been given. The highest average weight is again assigned to the speed domain. Yet, based on Figure 20 and Figure 21 it can be noted that the way of formulating the question leads to different results for some dimensions as clearly illustrated by the trauma management domain. In accordance with the analytic hierarchy process, we consider the average indicator weights (see also the third row of Table 17) in the computation of the index scores.

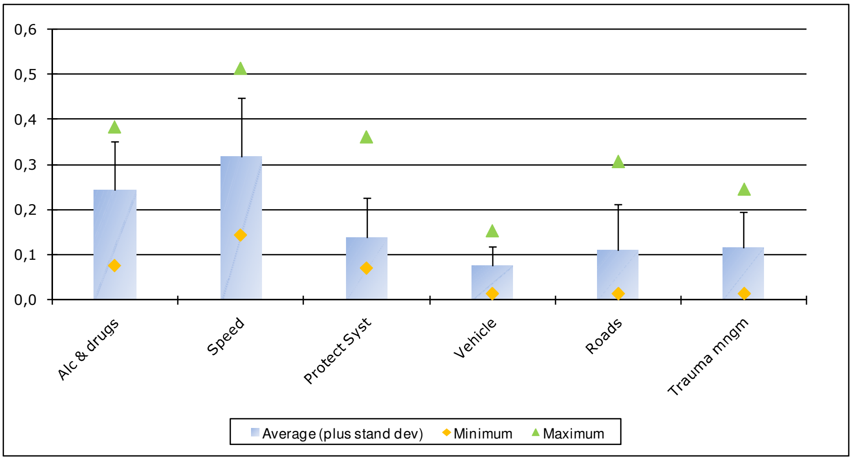


Figure 21: Budget allocation weights from a panel of experts

Data envelopment analysis is the fourth weighting method under study. Several software packages exist (see e.g., Barr, 2004). In this research, we opt for the LINGO software (LINDO Systems Inc., 2007) with its extensive options. We apply the data envelopment analysis model specified in (Eq. 6–2) to the data set (without normalisation), thereby maximising the sum of the 21 composite index scores while satisfying all imposed restrictions (see e.g., the maximum share of each indicator in the overall index score, presented in the first row of Table 16). Under the best set of weights the composite index value of three countries – Germany, Sweden and United Kingdom – equals one. For each country, the six indicator shares are given in Table 16 with the last column representing the composite index scores. The average DEA shares are shown in the fourth row of Table 17. It can be seen that the share of roads in the composite index is very limited[23]. Moreover, no other set of indicator weights could result in a higher sum of composite index scores. It appears that under the most favourable set of indicator weights Cyprus and Estonia are performing 16 respectively 11% less than the three best performing countries.

[23] The objective value (i.e., the sum of the 21 composite index scores) would not benefit from a higher weight assigned to the roads indicator because this would only increase the index score of the Netherlands and Belgium and be disadvantageous for the other countries having a relatively low roads score (see Figure 15).

Table 16: Indicator contributions and composite index values based on data envelopment analysis and (budget allocation) expert information

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM	CI SCORE
Maximum	*0.371*	*0.488*	*0.297*	*0.149*	*0.277*	*0.239*	
Austria	0.328	0.440	0.103	0.028	0.003	0.051	0.952
Belgium	0.317	0.412	0.088	0.032	0.008	0.064	0.922
Cyprus	0.263	0.412	0.107	0.013	0.004	0.044	0.843
Czech Rep	0.330	0.440	0.100	0.015	0.001	0.051	0.938
Denmark	0.336	0.449	0.112	0.031	0.004	0.061	0.993
Estonia	0.331	0.412	0.100	0.014	0.000	0.036	0.893
Finland	0.336	0.440	0.119	0.023	0.000	0.051	0.968
France	0.320	0.435	0.130	0.029	0.003	0.069	0.985
Germany	0.329	0.435	0.126	0.030	0.005	0.076	1.000
Greece	0.310	0.440	0.053	0.029	0.001	0.068	0.901
Hungary	0.332	0.412	0.079	0.021	0.001	0.057	0.902
Ireland	0.328	0.454	0.114	0.042	0.000	0.050	0.988
Italy	0.312	0.412	0.095	0.028	0.003	0.057	0.907
Netherlands	0.330	0.435	0.115	0.031	0.009	0.067	0.987
Poland	0.336	0.435	0.095	0.018	0.000	0.044	0.928
Portugal	0.323	0.416	0.118	0.023	0.003	0.066	0.948
Slovenia	0.328	0.440	0.108	0.040	0.003	0.060	0.979
Spain	0.313	0.416	0.115	0.029	0.003	0.053	0.928
Sweden	0.337	0.444	0.123	0.031	0.001	0.064	1.000
Switzerland	0.323	0.449	0.110	0.031	0.005	0.078	0.996
Unt Kingdom	0.335	0.449	0.124	0.035	0.002	0.055	1.000

Fifthly, the equal weighting method assigns to each indicator a weight of 1/6 or 0.167. These values are shown in the last row of Table 17.

Table 17: Four sets of indicator weights and the average indicator contribution (*) of data envelopment analysis

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
Factor Analysis	0.155	0.173	0.232	0.110	0.147	0.183
Analytic Hierarchy Process	0.248	0.287	0.134	0.098	0.173	0.060
Budget Allocation	0.243	0.319	0.138	0.075	0.111	0.115
Data Envelopment Analysis (*)	0.324	0.432	0.106	0.027	0.003	0.058
Equal Weighting	0.167	0.167	0.167	0.167	0.167	0.167

Of the five weighting methods, data envelopment analysis is the most different technique. For this method, instead of six indicator weights summing up to one, the average indicator shares are presented in Table 17. Comparing the weights from the other four methods shows the similarity between factor analysis and equal weighting on the one hand and analytic hierarchy process and budget allocation on the other. With respect to the equal weights and the weights based on factor analysis it can be seen that the variation in weights is rather limited. However, in case factor analysis is used instead of equal weighting, the protective systems, trauma management and speed indicator obtain a relatively higher weight while vehicle, roads and alcohol/drugs are assigned less weight. The weights resulting from the participatory methods AHP and BA starting from the same group of experts are more or less alike. In general, the experts label vehicle and trauma management as less influencing factors while the alcohol and drugs domain and the speed domain are considered to be the most important risk determinants of road safety. Moreover, based on accident reports, speeding and drink driving can be identified as the most contributing factors in fatal accidents in Europe (Organisation for Economic Co-operation and Development and European Conference of Ministers of Transport, 2006).

6.2.2 Rankings

Each set of weights is used to calculate the road safety index scores of the 21 countries. These scores result from the sum of the product of each (normalised) indicator value and its assigned weight[24]. Based on the index scores the countries are ranked in such a manner that a higher index score relates to a better position in the ranking. In other words, rank 1 corresponds to the country with the best (i.e., highest) index value. Table 18 shows the countries' rankings resulting from the five weighting methods and the final outcome ranking based on the number of road fatalities per million inhabitants.

First, we consider the five composite index rankings. For each country, the average rank over the five weighting methods can be computed based on Table 18. Germany, the Netherlands and Switzerland have the best rank (i.e., 3) followed by Denmark and United Kingdom. In fact, from the ordinal data set presented in Table 10 (page 108) it can be deduced that only these five countries have at least two top-five ranks and never rank worse than 15th on a risk domain. Overall, there is a high degree of correlation between the index rankings (see Table 19). Table 18 shows that the five weighting methods even completely agree on the ranking of Cyprus and Estonia. These countries have a very low index rank regardless of the weighting method used. The highest disagreement between the five options relates to Belgium, the Netherlands and Sweden (having a standard deviation larger than two). The discrepancy (mainly between DEA and the other methods) can be explained by the very large roads value of the Netherlands and Belgium which in case of DEA is only limitedly taken into account in the index. Sweden on the other hand only performing very bad with respect to the roads indicator gains on average five positions when considering the DEA ranking. Furthermore, Italy and Spain have their worst rank in case of BA weights whereas Switzerland and Poland benefit from the high contribution of the alcohol/drugs and speed indicator. Finally, France and Germany obtain their best rank under the FA ranking which is reasonable given that this method assigns the highest weight to the dimension for which they are the top-2 countries, i.e., protective systems.

[24] Here, index scores are obtained applying linear aggregation. In Chapter 8, the issue of aggregation and various options are discussed in detail.

Table 18: Countries' rankings based on the five weighting methods and the number of road fatalities per million inhabitants

	FA ranks	AHP ranks	BA ranks	DEA ranks	EW ranks	FO ranks
Austria	12	10	11	11	11	11
Belgium	13	13	15	16	10	12
Cyprus	21	21	21	21	21	17
Czech Rep	15	12	12	13	15	18
Denmark	5	2	2	5	4	6.5
Estonia	20	20	20	20	20	13.5
Finland	11	11	10	10	13	4
France	6	9	9	8	8	9
Germany	1	5	5	2	2	6.5
Greece	18	17	16	19	17	19
Hungary	19	19	18	18	19	16
Ireland	9	6	7	6	9	8
Italy	16	18	19	17	16	10
Netherlands	2	1	4	7	1	3
Poland	17	15	13	15	18	21
Portugal	10	14	14	12	12	20
Slovenia	8	8	8	9	6	13.5
Spain	14	16	17	14	14	15
Sweden	7	7	6	2	7	1
Switzerland	3	4	1	4	3	5
Unt Kingdom	4	3	3	2	5	2

Given the indicator set and the available information, a researcher can opt for one of the weighting methods described here. However, we also want to express which method is to be preferred based on the current data set. For the composite road safety performance index, the final outcome ranking is a relevant point of reference (see also section 2.9). The degree of correlation between each index ranking and the final outcome ranking is shown in Table 19. The five correlation coefficients lie between 0.72 and 0.77. The ranking resulting from the data envelopment analysis method has the best fit with the ranking based on the number of road fatalities per million inhabitants. In other words, given the data set the optimization algorithm of DEA best resemblances the final outcome ranking.

The results in Table 18 and Table 19 point out the harmony between the different rankings. Nevertheless, based on a qualitative assessment of the five weighting methods the data envelopment analysis method can be preferred. The best possible weights are derived taking the performances of all countries in the data set into account. At the same time, acceptable weights are obtained respecting the imposed upper limit on the share of each indicator in the index score. The model results in optimal index scores and offers clear insight into the underlying mechanisms and the best performing countries.

Table 19: Spearman's correlation coefficients between five index rankings based on a specific weighting method and the final outcome ranking

	FA RANKING	AHP RANKING	BA RANKING	DEA RANKING	EW RANKING	FO RANKING
FA RANKING	1.000	0.943	0.923	0.938	0.979	0.741
AHP RANKING		1.000	0.978	0.930	0.951	0.744
BA RANKING			1.000	0.943	0.914	0.718
DEA RANKING				1.000	0.900	0.769
EW RANKING					1.000	0.732
FO RANKING						1.000

Finally, a more detailed description of the main dissimilarities between the index rankings and the final outcome ranking (based on Table 18) is given. The first type of countries, i.e., Cyprus and Estonia, have an index rank which is very different from the final outcome rank regardless of the weighting method used. Therefore, better and probably more indicators are needed to approximate their position in the final outcome ranking. The second type of countries – Finland, Italy and Sweden – have a much better rank position in terms of road fatalities per million inhabitants than based on their index score. Moreover, the five weighting methods disagree in terms of their ranks. On average, DEA provides the closest link with the final outcome rank for this group of countries. As this method assigns very limited weight to the roads indicator, it might be an indication of the inappropriateness of the selected roads indicator for this category of countries. Thirdly, Czech Republic, Poland, Portugal and Slovenia obtain a better position based on the performance index than based on their number of fatalities per million inhabitants. On average, the best approximation with the final outcome rank occurs in case of equal weighting, where compared to the other weighting methods, a small share of the weight is devoted to the alcohol and drugs indicator and the speed indicator. As Czech Republic and Poland score best on these two dimensions their index score is lower in case of equal weights, bringing their position in the ranking closer to the final outcome ranking. In other words, for these countries the alcohol/drugs and speed performance has a rather unclear relationship with the final road safety outcome. Finally, for 9 of the 21 countries the absolute difference in rank between the average of the five index rankings and the final outcome ranking is limited (i.e., a maximum of two positions).

6.3 Concluding remarks

In this chapter, five common methods for assigning weights to indicators have been discussed. The process of deducing indicator weights based on factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting was described and attention has been paid to the advantages, disadvantages and requirements for each method. Different methods are appropriate under different circumstances. A table with summary information was given to help in selecting a justifiable weighting method.

Apart from the theoretical considerations each method was applied to the road safety indicator data set. Each method resulted in six indicator weights thereby implying a relatively higher or lower contribution of an indicator to the overall index (e.g., experts stressed the importance of alcohol/drugs and speed). Consequently, different countries' rankings were obtained. In general, high degrees of correlation were found between the five index rankings.

In the end, we want to develop a scientifically sound and appropriate methodology for the creation of a composite road safety performance index. The selection of the weighting method is an importance aspect herein. Of the five methods, the data envelopment analysis method resulted in a ranking which best approaches the final outcome ranking based on the number of road fatalities per million inhabitants. Apart from the positive quantitative evaluation, this chapter showed the value of DEA in the road safety index context. By formulating an optimization problem, as in (Eq. 6–2), best possible and at the same time acceptable indicator weights could be obtained. The DEA model strongly weights the behavioural indicators – alcohol and drugs, speed and protective systems – and assigns very limited weight to the roads indicator.

In the next chapter, the data envelopment analysis method is extended. More specifically, input-output optimization models are created based on which relevant benchmarks can be determined and road safety enhancing recommendations can be made for underperforming countries.

Chapter 7 Optimization method for prioritising risk factors and identifying benchmarks[25]

This chapter does not deal with a specific methodological step in the index process but further elaborates the data envelopment analysis technique described in Section 6.1.4. Both safety performance indicator data (e.g., seat belt wearing rates) and final outcome indicator data (e.g., the number of road fatalities per million inhabitants) are simultaneously studied here. In addition, country-specific optimization models (rather than one general model) will be formulated. Based on the model output, the good and bad road safety aspects of each country are identified. Moreover, valuable policy supporting information results from this approach. Rather than postulating the same set of benchmarks – e.g., the SUNcountries Sweden, United Kingdom and the Netherlands – and the same measures for each country, this methodology takes the characteristics of each country into account. For an underperforming country, another country in the data set will be assigned as useful benchmark. Finally, targets can be set and the risk factors prioritised thereby gaining a clear understanding of the type of policy actions most urgently needed.

This chapter consists of four sections. First, the new DEA model is discussed (Section 7.1). Second, the results are given in Section 7.2. Section 7.3 discusses the advantages and limitations of the model. Finally, the main conclusions are summarised in Section 7.4.

7.1 Methodology

As stated in the previous chapter, data envelopment analysis is a performance measurement technique that can be used for evaluating the relative efficiency of countries. As opposed to the DEA model described in the previous chapter, both inputs and outputs will be considered. More specifically, similar to the original DEA context, the

[25] Related research has been published in: Hermans, E., Brijs, T., Wets, G. and Vanhoof, K. (2009a). Benchmarking road safety: Lessons to learn from a data envelopment analysis. Accident Analysis and Prevention, 41 (1) pp. 174-182.

ratio of the weighted sum of outputs to the weighted sum of inputs is optimized (Cooper et al., 2004). A set of best weights is determined resulting in the most optimal objective value for a country while taking into account a particular set of road safety inputs and outputs. For the road safety case, the number of injury accidents and the number of fatalities are possible final outputs while the performance on the underlying risk domains can be seen as the inputs. By defining output and input in this way, the logical relationship of inputs leading to outputs is preserved. For example, an increase in the seat belt wearing rate results in a reduced number of fatalities.

However, as opposed to the economics field, we aim for a road safety output that is as low as possible and performance indicators that are as high as possible. Therefore, the ratio of the weighted sum of outputs and the weighted sum of inputs will be minimized. A linear model is formulated in which the sum over k weighted output values of a country j is minimized and the sum over l weighted safety performance indicator values of that country is set equal to one. Algebraically, the DEA model that will be used in this chapter is presented in (Eq. 7–1).

For each country in the data set $(j = 1,\ldots,n)$ the most optimal road safety score under a number of restrictions is determined. A country obtaining a road safety score of one is a best performing country; underperforming countries on the other hand have a road safety score larger than one. The reasoning is that a certain amount of safety performance (input) results in some level of accidents and fatalities (output). In case the output in terms of accidents and fatalities is higher than what could be expected based on the safety performance (in other words, weighted sum of outputs minus weighted sum of inputs is larger than zero), the country is labelled as underperforming and its road safety score will be larger than the optimal minimum of one. So, the better performing country of two countries with the same level of performance (i.e., the same performance indicator values) is the one with the lowest number of accidents and fatalities. The first inequality restriction in (Eq. 7–1) guarantees that the difference between the weighted sum of outputs and the weighted sum of inputs is non-negative for all n countries in the data set in case the best weights for the country under study (i.e., j) are filled in. In other words, no country can have a weighted sum of outputs smaller than its weighted sum of inputs.

$$RSS_j = \min \sum_{o=1}^{k} w_{oj} y_{oj} \qquad \text{(Eq. 7–1)}$$

$$\text{s.t.} \quad \sum_{i=1}^{l} v_{ij} x_{ij} = 1$$

$$\sum_{o=1}^{k} w_{oj} y_{om} - \sum_{i=1}^{l} v_{ij} x_{im} \geq 0 \qquad \forall m = 1,\ldots,n$$

$$L_o \leq \frac{w_{oj} y_{oj}}{\sum_{o=1}^{k} w_{oj} y_{oj}} \leq U_o \qquad \forall o = 1,\ldots,k$$

$$l_i \leq v_{ij} x_{ij} \leq u_i \qquad \forall i = 1,\ldots,l$$

$$w_{oj}, v_{ij} \geq 0 \qquad \forall o, i$$

with RSS = road safety score

j = country under study ; m = one of the countries in the data set

k = number of outputs ; l = number of inputs

w_{oj} = weight of output o of country j ; v_{ij} = weight of input i of country j

y_{oj} = value of output o of country j ; x_{ij} = value of input i of country j

L_o = lower output limit ; l_i = lower input limit

U_o = upper output limit ; u_i = upper input limit

The final restriction guarantees non-negative input and output weights. Furthermore, additional restrictions are incorporated to take prior knowledge or accepted views with respect to the input and output shares into account (Allen et al., 1997; Pedraja-Chaparro et al., 1997). Otherwise, a country with relatively few accidents, an average number of fatalities and a very good speed performance would get assigned all output weight to the accident dimension and all input weight to the speed dimension. In that case, only two aspects of road safety (i.e., accidents and speed) are considered which is unacceptable for fairly comparing countries. Therefore, restrictions for the input and output weights were added.

The DEA model will be illustrated using six road safety risk dimensions and two final outcome dimensions (as presented in Figure 22). The 21 countries focused on in the other chapters of this manuscript are used here as well. Moreover, each dimension is represented by one indicator and 2003 data will be used. The six best available safety performance indicators are the inputs. On the output side, accident and fatality data are used. As discussed earlier (see Section 4.3) the most common final outcome indicator uses the number of road fatalities normalised by data on the population size. In addition to the severity of accidents the frequency of injury accidents is considered here. More specifically, the two output indicators are the number of road fatalities per million inhabitants and the number of injury accidents per 100,000 inhabitants. Data are deduced from the European Union Road Federation (2007) and the Economic Commission for Europe (2008).

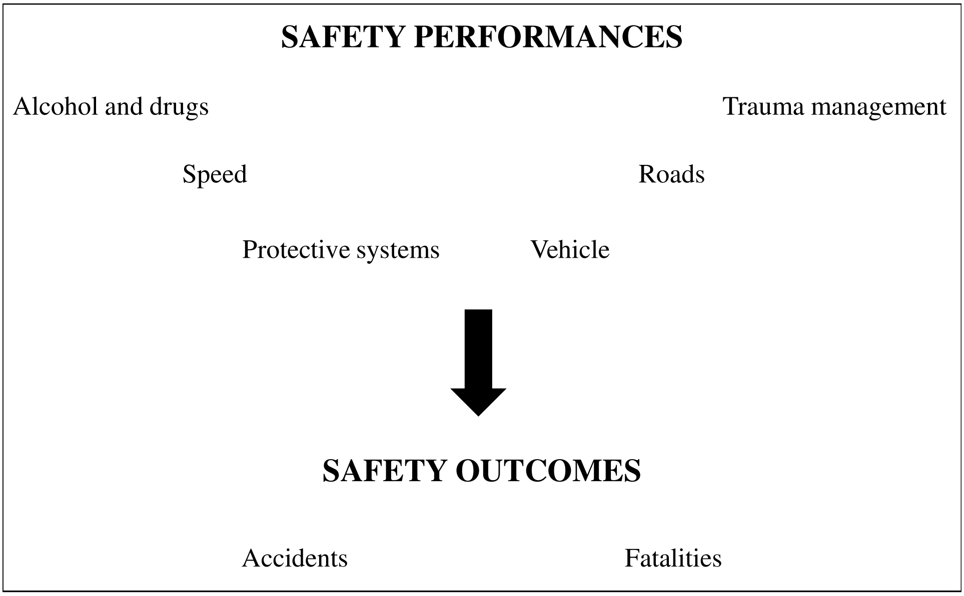


Figure 22: Road safety dimensions used in the input-output DEA model

The share of each input (i.e., the product of the performance indicator value and its weight) in the overall weighted sum of inputs is constrained to lie in a range defined by experts. In accordance with the DEA model presented in the previous chapter, the average of the two highest weights of eleven budget allocating road safety experts is used as upper limit for each performance indicator. Analogously, a lower limit is added[26]. All experts assigned some budget (i.e., more than €0) to the six domains. Consequently, all six safety performance indicators will be considered in the model. More specifically, with respect to the inputs the upper (*u*) and lower (*l*) limits presented in Table 20 need to be fulfilled.

Table 20: Lower and upper limits for the share of each indicator in the overall input

	LOWER LIMIT	UPPER LIMIT
ALC & DRUGS	0.077	0.371
SPEED	0.149	0.488
PROTECT SYST	0.072	0.297
VEHICLE	0.022	0.149
ROADS	0.015	0.277
TRAUMA MNGM	0.022	0.239

In addition, to guarantee that both output indicators will be used to some extent by the model, limits (*L* and *U*) are incorporated for the share of injury accidents and fatalities in the weighted sum of outputs. The accident share is restricted to lie within [0.1;0.5] while the interval for the fatality share is [0.4;0.8]. The two intervals are rather broad to allow a high level of flexibility. Moreover, the intervals overlap to enable each safety outcome to get assigned the largest share. However, the interval of the fatality share is higher than the one of the accident share due to the quality of these data (the definition of a road fatality is rather uniform across countries, these figures are well (i.e., accurately) collected and are assumed to be the correct ones). In addition, two risk domains, namely protective systems and trauma management are assumed to mainly affect the severity level of an injury and the probability of surviving an accident. Therefore, the fatalities dimension has a higher share than the accidents dimension.

[26] As opposed to the DEA model in Chapter 6, a feasible solution is now obtained respecting both upper and lower limits. This is due to the fact that each country has a separate optimization model and its own set of weights thereby implying a higher degree of flexibility to meet all restrictions.

Rather than comparing countries on each road safety aspect separately or per category (inputs versus outputs), one of the goals is to determine the most optimal road safety score for each country that considers all dimensions presented in Figure 22. The simultaneous study of safety outcomes and safety performances offers new insights. In general, the following results can be obtained. An overall ranking of the countries can be made based on their road safety score. Next, for underperforming countries with an index score larger than one, problem areas can be identified and at least one other country in the data set appointed as benchmark (this is the country for which the weighted sum of outputs is equal to the weighted sum of inputs). Based on the indicator values of the benchmark and a country-specific adjustment factor, useful targets can be set for the underperforming country and the achievement towards these targets could be monitored in the future. By comparing the current values with the target values the domains requiring urgent action can be identified and priorities assigned. These aspects will be illustrated in Section 7.2.

7.2 Results

Using road safety data of six inputs and two outputs for 21 European countries, the DEA model – algebraically presented in (Eq. 7–1) – yields the following results: a countries' ranking based on the road safety scores, an identification of relevant benchmarks for each underperforming country, detailed outcomes per country and country-specific road safety priorities. Each aspect is subsequently discussed.

7.2.1 Ranking based on the road safety scores

In Table 21, the 21 European countries are shown in increasing order of road safety score. For each country, the DEA model selects the most favourable weights that both satisfy the imposed restrictions and minimize the road safety score. A combination of low final outcome indicators and a number of high performance indicators can result in an optimal score of one in case all restrictions are met. The minimum score of one is obtained by two countries: Denmark and the Netherlands. One characteristic of a road safety score equal to one is that several sets of input and output weights result in this optimal score. Therefore, in the remainder of this section, we focus on underperforming countries with a road safety score larger than one and their unique, best weights.

Table 21: Countries' ranking based on road safety scores

	Road safety score
Denmark	1.000
Netherlands	1.000
Unt Kingdom	1.094
Sweden	1.133
France	1.183
Ireland	1.207
Switzerland	1.226
Finland	1.233
Germany	1.360
Greece	1.519
Estonia	1.646
Spain	1.804
Hungary	1.831
Italy	1.929
Belgium	2.039
Poland	2.073
Czech Rep	2.077
Austria	2.088
Slovenia	2.111
Cyprus	2.241
Portugal	2.349

7.2.2 Identification of relevant benchmarks for target setting

As shown by Table 21, 19 European countries cannot obtain a road safety score equal to one. This is due to a certain binding restriction. More specifically, the best weights for the country under study cause the weighted sum of outputs of (at least) one other country in the data set to become equal to its weighted sum of inputs. This country can then be seen as a realistic and valuable benchmark. The specific road safety characteristics of the country under study are taken into account in the selection of the benchmarks, i.e., the best weights stress the road safety dimensions a country performs on relatively well. Consequently, the benchmarks differ between the countries requiring improvement. The benchmark is a country, on the one hand, quite similar to the one under study and, on the other hand, a country that scores better than the country under study on most aspects. For example, Estonia should take Denmark as an example because with respect to output they both perform well in terms of the number of injury accidents per 100,000 inhabitants[27] (4th respectively 1st); in terms of input, they both perform best in terms of alcohol and drugs (7th respectively 3rd).

It is also possible that for more than one country in the data set the weighted sum of outputs becomes equal to its weighted sum of inputs. This is for example the case for Belgium and United Kingdom whereupon a combination of two benchmarks – the Netherlands and Sweden – should be taken as an example. Based on the model output, the share of each benchmark in the overall combination can be noted. Again, the road safety aspects of the country under study are taken into account. Comparing the safety performances of the two benchmarks (see Table 10) demonstrates that the Netherlands scores better with respect to road and trauma management whereas Sweden has better alcohol and drugs, speed and protective systems performances. Moreover, Belgium has a good vehicle, roads and trauma management performance and United Kingdom has excellent behavioural scores. As a result, Belgium should take a hypothetical country as an example which consists mainly of the Netherlands. For United Kingdom, Sweden takes the largest share in the combination. Table 22 indicates the (combination of) benchmark(s) for each of the 19 underperforming countries.

[27] Here, we assume a comparable degree of injury accident reporting in the 21 countries

Table 22: Benchmarks for underperforming countries

	Denmark	Finland	Netherlands	Sweden
Austria			X (28%)	X (69%)
Belgium			X (93%)	X (3%)
Cyprus	X (6%)	X (33%)	X (50%)	
Czech Rep	X (67%)		X (23%)	
Estonia	X (79%)			
Finland			X (70%)	
France	X (105%)			
Germany			X (56%)	X (49%)
Greece	X (93%)			
Hungary	X (88%)			
Ireland			X (81%)	
Italy			X (32%)	X (60%)
Poland			X (52%)	
Portugal	X (14%)		X (82%)	
Slovenia			X (35%)	X (67%)
Spain	X (71%)	X (4%)	X (16%)	X (4%)
Sweden			X (85%)	
Switzerland			X (51%)	X (54%)
Unt Kingdom			X (19%)	X (83%)

Moreover, the identification of the benchmark(s) and the percentages shown in Table 22 provide useful information in terms of target setting. (Eq. 7–2) illustrates the values an underperforming country *A* – with a road safety score larger than one and with *B* benchmarks – should aim at:

$$\text{Targets}_A = \sum_{b=1}^{B} \left(\frac{|\text{dual price}_b|}{RSS_A} \times \text{values}_b \right) \qquad \text{(Eq. 7–2)}$$

The target values of an underperforming country can be obtained based on the values of its benchmark(s) and a country-specific adjustment factor[28]. The latter factor – corresponding to the percentages in Table 22 – implies that underperforming countries with the same benchmark(s) have different target values. For example, the data of the Netherlands are used in the target setting of Sweden as well as Poland. However, Sweden has higher values to aim at than Poland (i.e., 85% respectively 52% of the values of the Netherlands), partly due to the fact that it has a lower road safety score.

In case of several benchmarks the adjustment factor acts as a weighting factor for the values of the benchmarks. The target values for Belgium are the sum of the values of the Netherlands multiplied by 93% and the values of Sweden multiplied by 3%. Finally, multiplying the targets obtained by means of (Eq. 7–2) with the best weights for that country results in a road safety score of one for the country under study.

7.2.3 Detailed model outcomes

For each country, the data envelopment analysis model in (Eq. 7–1) resulted in the most optimal road safety score under the imposed equality and inequality restrictions. The model output can be interpreted and translated in road safety enhancing recommendations. The approach will be illustrated here for the first two countries in the data set, being Austria and Belgium. In Section 7.2.4, an overview of the road safety risk domains requiring extra attention is given for all underperforming countries.

The results of the input-output data envelopment analysis model for Austria are presented in Table 23. For each input and output the final weight is given. These weights should not be compared with each other because raw values were used, having a different scale. It is better to look at the shares which are the products of the weight and the corresponding input or output value. The input restriction of the model implies that the sum of the six indicator shares equals one. The speed indicator contributes most to the overall input as it is the best of the six safety performance indicators for Austria. The sum of the accident share

[28] The adjustment factor is the ratio of the dual price of the binding restriction 'weighted sum of outputs equal to weighted sum of inputs' and the road safety score; the dual price is the rate at which the objective value will improve if the right-hand side or constant term of the restriction is increased by a small amount (LINDO Systems Inc, 2007); this information is given in the model output.

and the fatality share is equal to the road safety score of the country. For Austria, the final score is 2.088. This, and not one, is its most optimal score under the imposed restrictions.

Another piece of information is related to the input and output restrictions. There are binding restrictions for four inputs and one output. As we aim to minimize the road safety output (in terms of accidents and fatalities), more weight is given to the output dimension Austria scores on better. In this case, the fatality share makes up 80% of the output. This is due to the fact that of the 21 countries in the data set Austria has an average number of road fatalities per million inhabitants (rank 11) and a very high number of injury accidents per 100,000 inhabitants (rank 20). The weighted sum of inputs (which has to be equal to one) needs to consider all six performance indicators to some extent. The minimum and maximum indicator shares were determined taking the opinion of road safety experts into account and were presented in Table 20. From Table 23 it can be deduced that for the speed domain the maximum share is reached whereas the minimum share is binding for the protective systems, vehicle and trauma management domain. This already implies that some policy attention is needed for the latter three aspects of road safety.

Table 23: Detailed data envelopment analysis outcomes for Austria

	WEIGHTS	SHARE	BINDING RESTRICTIONS	TARGET VALUE	CURRENT VALUE	% CHANGE NEEDED
Alc & drugs	0.0038	0.371		96.42	97.4	-1.0
Speed	0.0052	0.488	Upper	91.54	94	-2.6
Protect syst	0.0009	0.072	Lower	87.52	77	13.7 (2)
Vehicle	0.0006	0.022	Lower	38.08	35.14	8.4
Roads	0.0126	0.025		1.98	2	-1.0
Trauma mngm	0.0029	0.022	Lower	9.23	7.5	23.0 (1)
Accidents	0.0008	0.418		195.68	535	
Fatalities	0.0145	1.671	Upper	58.32	115	
Road safety score		2.088				
Benchmarks (dual prices)			NL (-0.583) & SE (-1.441)			

The results for Austria give some useful insight. However, information from other, best performing countries is also available. Using the set of weights given in the second column of Table 23, the weighted sum of outputs for both the Netherlands and Sweden equals its weighted sum of inputs; this is a binding restriction. Therefore, a combination of the Netherlands and Sweden can be taken as an example. The dual price associated with the binding restriction for the Netherlands (-0.583) and for Sweden (-1.441) offers valuable information in this respect. Targets for Austria will be specified using the values of the benchmarks and the corresponding adjustment factors (Eq. 7–2). For Austria, the input and output values of the Netherlands are multiplied by 0.279 (i.e., 0.583/2.088) and the input and output values of Sweden by 0.690 (i.e., 1.441/2.088). The target values for Austria result from the sum of these two factors. Multiplying these target values in column 5 with the corresponding weights in column 2 results in a road safety score equal to one.

Furthermore, it is useful to compare the target values with the current Austrian situation and quantify the effort needed to attain the goals in terms of percentages. As measures should be taken on the input side, thereby affecting the output in terms of accidents and fatalities, we focus on the safety performance indicators here. Priorities for the road safety risk domains can be set based on the information in the last column of Table 23. It appears that Austria should try to improve its performance with respect to trauma management and protective systems. Furthermore, the vehicle domain requires an increase of 8.5% to aim at compared to the current performance. Finally, a negative value of the change needed for an indicator (in this case the alcohol and drugs, speed and roads domain) indicates that the current value exceeds the target value. Given the data set, Austria would be a best performing country in case it had 196 injury accidents per 100,000 inhabitants and 58 road fatalities per million inhabitants, a better road safety performance concerning protective systems, vehicle and trauma management (i.e., +13.7%; +8.4% respectively +23%) and a slightly reduced alcohol/drugs, speed and roads performance (-1%; -2.6% respectively -1%).

In a similar way, the outcomes for Belgium could be studied in order to get insight into its current and target performance on the different road safety aspects. Table 24 demonstrates that the most optimal road safety score for Belgium under the developed DEA model equals 2.039. In this score, the fatality share is much more represented than the accident share indicating the relatively better performance with respect to fatalities. Concerning the safety performance indicators, the vehicle and trauma management domain are stressed while the share of the speed and protective systems indicator equals the lower limit. The benchmarks appear to be the Netherlands and Sweden again but their dual prices indicate that for Belgium the target values are mainly affected by the values of its neighbouring country, the Netherlands. A comparison between the current and the target input values reveals that Belgium should aim at a large increase in the percentage of persons wearing their seat belt in traffic. Higher enforcement on the one hand and directed sensibility campaigns on the other are effective measures in this respect (SUPREME, 2007). Belgium would belong to the set of best performing countries in the data set in case the target values were met.

Table 24: Detailed data envelopment analysis outcomes for Belgium

	WEIGHTS	SHARE	BINDING RESTRICTIONS	TARGET VALUE	CURRENT VALUE	% CHANGE NEEDED
Alc & drugs	0.0036	0.337		93.49	94.2	-0.8
Speed	0.0017	0.149	Lower	88.64	88	0.7
Protect syst	0.0011	0.072	Lower	82.08	66	24.4 (1)
Vehicle	0.0036	0.149	Upper	37.08	40.84	-9.2
Roads	0.0095	0.054		5.66	5.7	-0.8
Trauma mngm	0.0254	0.239	Upper	9.32	9.4	-0.8
Accidents	0.0009	0.408		185.08	458	
Fatalities	0.0139	1.631	Upper	59.90	117	
Road safety score		2.039				
Benchmarks (dual prices)			NL (-1.888) & SE (-0.055)			

7.2.4 Road safety priorities per country

For all countries with a road safety score larger than one the approach illustrated in the previous section can be followed to reveal the risk domains that each underperforming country should work on. In Table 25, the most urgent road safety performance aspect is assigned score 1, the second most important aspect score 2 and so on. The table only mentions a score if the target value (based on the benchmarks' indicator values and the adjustment factors) is 10% higher than the current value.

Like Belgium, Italy, Slovenia and Switzerland should focus their road safety efforts in the first place on (frontal) seat belt wearing behaviour. Cyprus could work on the promotion of new(er) cars. Trauma management is the major bottleneck in Austria and United Kingdom while the Czech Republic, Estonia, Finland, France, Greece, Hungary, Ireland, Poland, Portugal, Spain and Sweden could gain from policy actions inspired on the roads domain[29]. Eight countries have one main priority (i.e., for one risk domain their current situation is more than 10% worse than what they should aim at) while five countries (Cyprus, Czech Republic, Estonia, France and Hungary) are urged to focus on at least three of the six dimensions. Finally, for Germany the target values do not significantly differ from the current values on any of the risk dimensions. From e.g., Table 8 it can be seen that Germany scores average to well on all safety performance indicators[30]. Comparing its current and target input values the vehicle domain requires the highest change (i.e., +8.1%).

It is difficult to validate the countries' priorities as very few related studies exist in literature. Here, we briefly mention the country reports on road safety performance (Organisation for Economic Co-operation and Development and European Conference of Ministers of Transport, 2006) in which the major current road safety problems indicated by each country are reported. Although half of the countries point out speed as the main problem and only two countries explicitly mention trauma management, some similarities can be found (e.g., the protective

[29] This can be concluded based on the (best available) indicators used. However, it can be assumed that the priority list for some countries changes in case other safety performance indicators are used. Due to the excellent performance of the Netherlands in terms of road safety output and its good performance on all risk domains, it is a frequent benchmark. Given its extremely high roads value, this domain often turns out to be the largest problem area. In the interpretation of the results the indicators used should be kept in mind.

[30] Nevertheless, it does not have a road safety score close to one due to its 18th position on the number of injury accidents per 100,000 inhabitants.

systems problem in Austria and Belgium; and the roads problem in Finland and Greece).

Table 25: Road safety priorities per underperforming country

	ALC & DRUGS	SPEED	PROTECT SYST	VEHICLE	ROADS	TRAUMA MNGM
Austria			2			1
Belgium			1			
Cyprus	3.5			1	3.5	2
Czech Rep				2	1	3
Estonia				2	1	3
Finland					1	
France	3			2	1	
Germany						
Greece			2		1	
Hungary			3	2	1	
Ireland					1	
Italy			1			
Poland					1	
Portugal				2	1	
Slovenia			1			2
Spain					1	2
Sweden					1	
Switzerland			1			
Unt Kingdom						1

7.3 Advantages and limitations

Having developed and applied a road safety input-output data envelopment analysis model, the evaluation in terms of advantages and limitations is discussed now. Road safety is complex and several aspects (final outcomes and safety performance indicators among other things) can be considered when comparing countries based on road safety. Data envelopment analysis is a technique in which multiple inputs and outputs can be incorporated, making it a useful technique for the road safety context. Taking into account final outcome and intermediate outcome data for a large set of countries, the most optimal road safety score is determined for each country and best performing countries can be distinguished from underperforming ones. To guarantee consistency with prior knowledge and accepted views from experts, restrictions with regard to the share of each input and output can be added. Countries are ranked based on their road safety score and areas of underperformance identified.

Since each country has its own road safety characteristics, the identification of one or more country-specific benchmarks can be justified. Taking these countries as an example for improving performance, useful targets and priorities for policy action can be set. That way, risk domains that need urgent action can be tackled and relevant measures can be taken. This makes the proposed model a valuable tool for the road safety field.

However, like any technique, DEA is based on a number of assumptions and characterized by some limitations that should be taken into account when interpreting the results. First, DEA compares the performance of a country to the performance of the other countries in the data set (Anderson, 2006). A change in the sample of countries (e.g., adding an extra country) may imply other outcomes. Moreover, the results produced by DEA are particularly sensitive to measurement error, input and output specification and sample size (Steering Committee for the Review of Commonwealth/State Service Provision, 1997). Therefore, reliable data sources should be consulted, appropriate inputs and outputs selected and as much – comparable[31] – countries as possible considered to enable a fair performance evaluation. Results should be interpreted with respect to the data set used.

As shown above, valuable information can be obtained from this model. However, this is mainly the case for underperforming countries. On the contrary, for the set of best performing countries, several sets of weights result in a score of one. Therefore, not much attention should be given to the weights presented in the output. In addition, as these countries are the best performing countries in terms of road safety within the data set, it is currently impossible to choose between them, identify a useful

[31] The countries focused on need to have a similar development in mobility as the inputs (and outputs) selected here may not be relevant for less developed countries.

benchmark or set challenging targets based on the data of the other countries. In future research, adaptations to the model can be made in order to render this data envelopment analysis model of value for all countries. In case data are available over time, this information could serve for target and priority setting.

7.4 Concluding remarks

In order to further reduce the level of accidents and casualties, an efficient road safety policy is required. In this respect, policymakers need to deal with the main problem areas. In this chapter, a model has been developed able to assist in prioritising actions. Apart from the benefits for the road safety field in terms of policy supporting information, the data envelopment analysis model presented here offers new insights for the composite index field. First, besides performance indicator values, accident and fatality data are used. The model results in the optimal road safety score for each country based on six inputs and two outputs, thereby respecting the opinion of experts regarding the share of e.g., speed in the overall input score. Secondly, we defined a minimization problem in which the performance indicator values (the input) should be as high as possible while the output (i.e., accidents and fatalities) should be minimized.

The model has a number of limitations that should be kept in mind. The country-specific results depend on the countries in the data set and the conclusions are affected by the inputs and outputs used in the model. Furthermore, for countries obtaining the optimal road safety score of one, no policy recommendations can be made at this time. For each underperforming country, based on its benchmark values and some adjustment factor, target values were formulated and subsequently compared to its current indicator values. The risk domains needing action were identified.

In the future, more aspects should be investigated. First, more inputs and outputs could be used to describe road safety, for example helmet wearing rates and the number of seriously injured persons. At the same time, the number of countries considered by the model could be increased. Secondly, this model is valuable for a disaggregated analysis as well in which relevant road safety inputs and outputs are compared for certain age classes (e.g., young persons), transport modes (e.g., motorcyclists) or on a more regional level. Thirdly, it would be valuable to take information about the cost-effectiveness of investments in different inputs into account. Fourthly, a sensitivity analysis could reveal the impact of a change in the weight restrictions or the impact of incorporating another indicator. Finally, data envelopment analysis is suitable for country comparisons over time as well.

Chapter 8 Aggregating indicators[32]

In this chapter, we focus on the sixth step in the index methodology, i.e., the aggregation of indicators. The mathematical formula for combining indicators needs to be selected. Various types of methods aggregating the different values in a single score exist in literature. In this respect, it is essential that the way of aggregating corresponds to the idea of the index. Special attention will be given to the class of ordered weighted averaging (OWA) operators as this comprehensible type of operators has interesting characteristics for the road safety index context.

In Section 8.1, a classification and description of common aggregation operators is given. In Section 8.2, OWA operators are discussed in detail. An algebraic elaboration is the topic of Section 8.3. In addition to some mathematical insights, the possibility of incorporating knowledge or opinions in the aggregation process is discussed in Section 8.4. The results of applying some OWA operators to the road safety indicator data set are discussed in Section 8.5. This chapter closes with a section (8.6) stating concluding remarks.

8.1 General concepts

Aggregation refers to the process of combining values in a single score such that the final result takes all individual values into account in a specific way. Aggregation is a very extensive research field in which numerous types of aggregation functions or operators exist. They are all characterized by certain mathematical properties and aggregate in a different manner. Aggregation operators can be used to express the importance of certain criteria (indicators in our case), to reflect the behaviour of the decision maker, to show interactions between the criteria, etc (Grabisch et al., 1999). Aggregation functions are for example used in case of multiple attribute decision making, group decision making and fuzzy logic (Beliakov et al., 2007). Here, the six safety

[32] Related research has been published in: Hermans, E., Ruan, D., Brijs, T., Wets, G. and Vanhoof, K. (2008a). *Evaluation of road safety performance indicators using OWA operators.* Presented at the 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid.

performance indicator values of each country will be aggregated in one index score based on which a countries' ranking can be set up.

8.1.1 Main classes of aggregation operators

In general, aggregation operators can be roughly divided into four broad classes: averaging functions, conjunctive functions, disjunctive functions and mixed functions (Beliakov et al., 2007). $\boldsymbol{x}$ ($x_1,x_2,...,x_n$) being a n-dimensional real vector, these aggregation classes can be defined as follows:

- An aggregation function f has averaging behaviour (or is averaging) if for every $\boldsymbol{x}$ it is bounded by $\min(\boldsymbol{x}) <= f(\boldsymbol{x}) <= \max(\boldsymbol{x})$.

- An aggregation function f has conjunctive behaviour (or is conjunctive) if for every $\boldsymbol{x}$ it is bounded by $f(\boldsymbol{x}) <= \min(\boldsymbol{x})$.

- An aggregation function f has disjunctive behaviour (or is disjunctive) if for every $\boldsymbol{x}$ it is bounded by $f(\boldsymbol{x}) >= \max(\boldsymbol{x})$.

- An aggregation function f is mixed if it does not belong to any of the above classes, i.e., it exhibits different types of behaviour on different parts of the domain.

Averaging is the most common way to combine scores. The basic idea is that the aggregated score cannot be higher or lower than any individual score. The aggregated value is a kind of representative value of all the individual values. The minimum (min) and maximum (max) functions can be considered as averaging; however, they are the limiting cases, on the border with conjunctive respectively disjunctive functions. In addition, conjunctive aggregation functions model conjunction (i.e., the logical AND) and disjunctive aggregation functions model disjunction (logical OR). Conjunction does not allow low scores for some criteria to be compensated by other scores. Consequently, the smallest individual value bounds the aggregated value. The opposite is true for disjunctive aggregation functions, i.e., satisfaction of any of the criteria is enough by itself, although the total score increases with more than one positive criterion. The prototypical examples of conjunctive and disjunctive aggregation functions are so-called triangular norms respectively conorms (t-norms and t-conorms) (Beliakov et al., 2007).

Numerous aggregation operators exist. They belong to one of the classes stated above and are grouped in families. There are no general guidelines for selecting a suitable aggregation operator. The researcher should try to make an educated choice based on the properties of the aggregation procedure and the context. Here, we focus on the class of averaging operators in which the index score is bounded by the lowest and highest indicator value.

8.1.2 Averaging operators

Two main types of averaging operators will be discussed, i.e., weighted mean operators and ordered weighted averaging operators.[33]

8.1.2.1 Weighted mean operators

Weighted mean operators are probably the best-known type of averaging operators. In fact, in the index construction context, the weighted arithmetic mean – referred to as linear aggregation – and the weighted geometric mean – referred to as geometric aggregation – are nowadays often used (see e.g., Saisana and Tarantola, 2002). As shown in (Eq. 8–1), in case of linear aggregation the index consists of the sum of each of the I indicator values (x_i) multiplied by its weight (w_i). Geometric aggregation, presented by (Eq. 8–2), results in an index score by raising each indicator value to the power of the corresponding weight and multiplying these products.

Given a weighting vector w (with $w_i \in [0,1]$ and $\sum_{i=1}^{I} w_i = 1$):

the weighted arithmetic mean = $\sum_{i=1}^{I} w_i x_i$ (Eq. 8–1)

the weighted geometric mean = $\prod_{i=1}^{I} x_i^{w_i}$ (Eq. 8–2)

[33] Other types of averaging operators exist (see e.g., Beliakov et al., 2007); for instance, Choquet and Sugeno integrals might have some value for the road safety index because interactions between criteria can be taken into account. Nevertheless, the high number of parameters (given the limited number of countries in the data set) and the complexity of these operators cause a further exploration to be a topic for future research.

8.1.2.2 Ordered weighted averaging operators

OWA functions are a second type of common averaging aggregation functions. They have been introduced by Yager (1988) and have become very popular in the fuzzy sets community. It is important to note that the weights in weighted means and in OWA functions represent different aspects. In case of OWA a weight is no longer associated with the meaning of a particular criterion (or indicator) – such as the alcohol and drugs weight – but with its magnitude. More specifically, the components of $\boldsymbol{x}$ are arranged in non-increasing order $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$ and the aggregated score results from multiplying the sorted values with the corresponding weights and summing these factors. In other words, in weighted means, w_i reflects the importance (or reliability or contribution) of the i^{th} indicator whereas in OWA it reflects the importance of the i^{th} largest indicator (Beliakov et al., 2007). To indicate that the weights apply to ordered data $\vec{w}$ will be used.

Ordered weighted averaging operators are promising for the current context as good and bad road safety performances can be weighted differently, regardless of the meaning of the indicator. Moreover, the attitude of persons (such as decision makers) in terms of the allowed degree of compensation between good and bad scores can be reflected (Yager, 1996). The concept of linguistic quantifiers (such as 'a few' or 'most') is useful to semantically convey the aggregation policy. This will be illustrated in Section 8.4. By changing the OWA parameters, different opinions can be represented in the aggregation.

8.1.3 Study design

The main classes of aggregation functions have been identified. For the construction of the road safety performance index, an operator belonging to the averaging aggregation class will be selected. More specifically, the ordered weighted averaging operators are studied in more detail. In the next section, it is shown that some aggregation operators such as the arithmetic mean and the minimum are special cases of an OWA operator.

OWA operators with different weighting vectors will be applied to the road safety indicator data set. Recall that the ordered weighted averaging operators require ordered data. Rescaled indicator values will be used resulting in index scores lying between zero and one. For each country, a reordering of the values presented in Table 8 takes place; the best performance (i.e., the highest of the six indicator values) obtains position one, followed by the second best performance, etc. This implies that the first value no longer automatically represents the alcohol and drugs performance. The road safety index scores are computed by multiplying the six ordered indicator values of each country with a specific weighting vector. At this stage, each indicator is assumed to be of equal importance (rather than using some other weighting set derived in Chapter 6) because this enables us to clearly assess the impact of the aggregation operator (and its specific weighting vector) on the end result.

8.2 OWA operators

In this section, ordered weighted averaging operators and their weighting vectors are discussed in more detail. First, some special types of OWA operators are given.

8.2.1 Maximum, minimum and arithmetic mean

Very common aggregation operators are maximum, minimum and arithmetic mean. In fact, they are all special cases of OWA operators. Below, the weighting vector $\vec{w}$ of these operators is given.

- max: $\vec{w}=(1,0,\ldots,0)$ considers only the best performance. It represents the optimistic point of view.
- min: $\vec{w}=(0,0,\ldots,1)$ considers only the worst performance. It represents the pessimistic point of view.
- arithmetic mean: $\vec{w}=(1/n,1/n,\ldots,1/n)$ considers each performance equally.

As illustrated by the weighting vectors, the maximum and minimum operator assign all weight to one criterion. In case all criteria are assigned an equal weight, OWA becomes the arithmetic mean.

8.2.2 Determining OWA weighting vectors

In the previous section, three weighting vectors have been given. Every combination of weights lying between zero and one and summing up to one can be used to aggregate the ordered values. Here, we discuss some methods for obtaining relevant OWA weights. First, the concepts of 'orness' and 'dispersion' are explained.

The degree of orness is an important numerical characteristic of averaging aggregation functions. It corresponds to the degree of optimism of the decision maker or the maxness of the aggregation (Yager, 1997). For an OWA weighting vector the degree of orness is defined as shown in (Eq. 8–3). Values lie between zero and one with orness(max)=1; orness(min)=0; and orness(arithmetic mean)=0.5.

$$orness(\vec{w}) = \frac{1}{I-1}\sum_{i=1}^{I}(I-i)\overrightarrow{w_i} \qquad \text{(Eq. 8–3)}$$

In addition to the orness value, the weights entropy or dispersion is an important parameter in choosing weighting vectors of OWA operators. The weights dispersion – defined in (Eq. 8–4) – measures the degree to which all the information (i.e., all criteria) is used in the aggregation process (Beliakov et al., 2007).

$$disp(\vec{w}) = -\sum_{i=1}^{I}\overrightarrow{w_i}\log\overrightarrow{w_i} \quad (\text{with } 0\times\log 0 = 0) \qquad \text{(Eq. 8–4)}$$

The following methods can be used to obtain OWA weights (Beliakov et al., 2007):

- Methods based on data: the weights are determined by solving an optimization problem in which the error between the aggregated scores and the real (known) scores is minimized. The orness can be added as an additional restriction to the optimization problem.
- Methods based on a measure of dispersion: a vector of weights is chosen that maximizes the dispersion given a predefined degree of orness. The solution is called maximum entropy OWA.
- Methods based on weight generating functions: Yager (1996) proposed to use regular increasing monotone (RIM) quantifiers. RIM quantifiers are fuzzy linguistic quantifiers that express the concept of fuzzy majority. More specifically, the decision maker states a linguistic quantifier *Q* (such as for all; most; ...) which offers a fuzzy description of the portion of criteria required to be met by a good solution. If *Q* is a RIM quantifier, OWA weights can be obtained using (Eq. 8-5) and most commonly (Eq. 8–6).

$$\overrightarrow{w_i} = Q(i/I) - Q((i-1)/I) \text{ for } i = 1,\ldots,I \qquad \text{(Eq. 8–5)}$$

$$Q(r) = r^{\alpha} \text{ with } \alpha \geq 0 \qquad (\text{with } 0^0 = 0) \qquad \text{(Eq. 8–6)}$$

In case of six criteria (or indicators) the above formulas result in:

$$\begin{aligned}&\overrightarrow{w_1} = [1/6]^{\alpha};\ \overrightarrow{w_2} = [2/6]^{\alpha} - [1/6]^{\alpha};\ \overrightarrow{w_3} = [3/6]^{\alpha} - [2/6]^{\alpha};\\ &\overrightarrow{w_4} = [4/6]^{\alpha} - [3/6]^{\alpha};\ \overrightarrow{w_5} = [5/6]^{\alpha} - [4/6]^{\alpha};\ \overrightarrow{w_6} = 1 - [5/6]^{\alpha}\end{aligned} \qquad \text{(Eq. 8–7)}$$

The six road safety performance indicators will be aggregated by means of an ordered weighted averaging operator. The weighting vector to use will be determined based on linguistic quantifiers specifying the degree to which the final index score should consist of its underlying values (see Section 8.4). This additional information will be taken into account in the aggregation process.

In the next section, the α parameter will be discussed. In general, it can be stated that α indicates the tendency of the operator to assign more weight to either higher or lower scores (Yager and Kacprzyk, 1997). In terms of road safety, α represents the degree to which the occurrence of road fatalities depends on the magnitude of the six performances. For α equal to one, the number of road fatalities per million inhabitants is considered to result equally from good and bad performances. An α value larger (smaller) than one implies that the worst (best) performances affect the number of road fatalities more and therefore low (high) indicator values are emphasized in that case. For different values of α, the ordered weights as well as the corresponding orness value is given below. Based on (Eq. 8-5) and (Eq. 8–6) the orness in (Eq. 8-3) can be formulated in terms of α:

$$orness(\vec{w}) = \frac{1}{I-1}\sum_{i=1}^{I}(I-i)\overrightarrow{w_i} = \frac{1}{I-1}\sum_{i=1}^{I-1}\left(\frac{i}{I}\right)^{\alpha} \qquad \text{(Eq. 8–8)}$$

8.3 Algebraic elaboration

In this section, the effect of changing α is illustrated. In theory, α can take on any value in the interval zero to infinity. The resulting six weights as well as the orness value for eight possible values of α are given in Table 26.

Table 26 contains the three special cases of OWA operators discussed in Section 8.2.1. More specifically, the maximum operator corresponds to an α equal to zero; the minimum operator to a very large α (e.g., 1×10^{10}); and the arithmetic mean operator to an α equal to one. By computing the weights for eight different α values it can be clearly seen that the share of weight given to worse performances increases with α. For example, an α value of four would create an index which consists for over 50% of the worst indicator value of a country.

Furthermore, the aspect of compensation – i.e., the degree to which offsetting of bad values by good values is allowed – can be illustrated by means of the values in Table 26. Small values of α allow some degree of compensation whereas in case of a high α, bad scores are not compensated by good ones. The last row in Table 26 shows the degree of orness (Eq. 8–8) for different values of α. A non-decreasing weighting vector results in an orness between 0 and 0.5 whereas a non-increasing weighting vector has an orness lying between 0.5 and 1. In general, an orness above 0.5 points out a tolerant aggregation behaviour whereas an orness smaller than 0.5 indicates rather intolerant behaviour.

Table 26: Ordered weights for different values of α

	0.0	0.5	1	1.5	2	3	4	$1\text{x}10^{10}$
$\overrightarrow{w_1}$	1.00	0.408	0.167	0.068	0.028	0.005	0.001	0.00
$\overrightarrow{w_2}$	0.00	0.169	0.167	0.124	0.083	0.032	0.012	0.00
$\overrightarrow{w_3}$	0.00	0.130	0.167	0.161	0.139	0.088	0.050	0.00
$\overrightarrow{w_4}$	0.00	0.109	0.167	0.191	0.194	0.171	0.135	0.00
$\overrightarrow{w_5}$	0.00	0.096	0.167	0.216	0.250	0.282	0.285	0.00
$\overrightarrow{w_6}$	0.00	0.087	0.167	0.239	0.306	0.421	0.518	1.00
orness	1.00	0.684	0.500	0.384	0.306	0.208	0.151	0.00

To conclude, the higher α, the more emphasis on weaker performances (and the lower the final score); the less compensation allowed; and the lower the degree of orness. The weighting vector corresponding to a specific value of α will be multiplied with the ordered indicator values to compute index scores. As criterion for selecting among the various α values, the correlation coefficients between the resulting rankings and the ranking based on the number of road fatalities per million inhabitants can be compared. However, rather than merely mathematically deducing the best weighting vector for the current data set, the emphasis should be on selecting an α that appropriately represents the aggregation idea of a decision maker. Therefore, in the next section, linguistic formulations are transformed in order to obtain a smaller set of relevant values of α.

8.4 Linguistic formulations

In this section, the aggregation opinion of a group of road safety experts is expressed by means of linguistic guidelines. Subsequently, these formulations are transformed into restrictions for α. Various guidelines could be formulated; however, the probability of conflicting formulations and restrictions for α increases with the number of formulations. Therefore, an order of importance should be assigned to the guidelines; restrictions from a next guideline will only be considered in case all former restrictions could be met. From a panel discussion, the following principles were derived[34]:

- in case a country scores badly on more than a few indicators, its final road safety index score should be small
- in case a country scores badly on a few indicators, its final road safety index score should be between small and average

Starting from the idea that all domains are essential risk factors leading to accidents and casualties, the final index should take all six indicators into account to some minimal extent. However, the occurrence of one or a few bad performances may have a large effect on the number of road fatalities and should therefore result in a lower index score. Moreover, it should not be allowed to completely counterbalance bad scores by an excellent score.

The first step in transforming the guidelines into restrictions for α is to give a specific meaning to the concepts ‘badly’ (with respect to indicator performance), ‘a few’ (with respect to the number of indicators), ‘small’ and ‘average’ (with respect to the index score). The performance with respect to a specific indicator will be classified as ‘good’, ‘average’ or

[34] Another approach would have been to deduce principles from real road safety data.

‘bad’ (Yager and Kacprzyk, 1997). Here, score 1 is assigned to good; score 0.5 to average and score 0 to bad performances. Next, compared to ‘at least one’ or ‘half’ or ‘all’ indicators, ‘a few’ or ‘most’ or ‘almost all’ indicators might cause some vagueness. On a total of six indicators, ‘a few’ corresponds to two; ‘most’ to four and ‘almost all’ to five. Finally, the classification of index scores is accomplished by dividing the possible interval [0,1] into four categories. A ‘small’ index score is 0.25 at most, an ‘average’ index score corresponds to 0.5 whereas a ‘large’ index score is at least 0.75.

Having assigned a meaning to all concepts in the linguistic formulations, the next step is to translate the formulations numerically. By using (Eq. 8–7) restrictions for α can be deduced. This is illustrated for the two guidelines in (Eq. 8–9) respectively (Eq. 8–10).

$$
\begin{aligned}
& f_\alpha(1,1,1,0,0,0) \le 0.25 \\
& \Leftrightarrow \overrightarrow{w_1} + \overrightarrow{w_2} + \overrightarrow{w_3} \le 0.25 \\
& \Leftrightarrow \left(\frac{1}{6}\right)^\alpha + \left(\frac{2}{6}\right)^\alpha - \left(\frac{1}{6}\right)^\alpha + \left(\frac{3}{6}\right)^\alpha - \left(\frac{2}{6}\right)^\alpha \le 0.25 \\
& \Leftrightarrow \left(\frac{1}{2}\right)^\alpha \le 0.25 \\
& \Leftrightarrow \alpha \ge 2
\end{aligned}
\qquad \text{(Eq. 8–9)}
$$

$$
\begin{aligned}
& 0.25 < f_\alpha(1,1,1,0.5,0,0) < 0.5 \\
& \Leftrightarrow 0.25 < \overrightarrow{w_1} + \overrightarrow{w_2} + \overrightarrow{w_3} + 0.5 \times \overrightarrow{w_4} < 0.5 \\
& \Leftrightarrow 0.25 < \left(\frac{1}{6}\right)^\alpha + \left(\frac{2}{6}\right)^\alpha - \left(\frac{1}{6}\right)^\alpha + \left(\frac{3}{6}\right)^\alpha - \left(\frac{2}{6}\right)^\alpha + 0.5 \times \left[\left(\frac{4}{6}\right)^\alpha - \left(\frac{3}{6}\right)^\alpha\right] < 0.5 \\
& \Leftrightarrow 0.25 < 0.5 \times \left(\frac{3}{6}\right)^\alpha + 0.5 \times \left(\frac{4}{6}\right)^\alpha < 0.5 \\
& \Leftrightarrow 1.2946 < \alpha < 2.6526
\end{aligned}
\qquad \text{(Eq. 8–10)}
$$

Based on (Eq. 8–9) and (Eq. 8–10) we can conclude that α should be in the interval [2;2.65] to aggregate the six indicators in a way that is acceptable for the experts (thereby respecting both formulations). Hence, for α equal to 2;2.1;2.2;2.3;2.4;2.5;2.6 we compute the corresponding weighting vector using (Eq. 8–7) and multiply it with the ordered indicator values to compute the index scores for the 21 countries.

8.5 Discussion

In this chapter, different aggregation operators are described. The results in terms of correlation between a road safety index ranking (based on a specific operator and weighting vector) and the final outcome ranking (using the number of road fatalities per million inhabitants) are presented in Table 27. Moreover, a brief qualitative discussion of each operator is given.

The maximum, minimum and arithmetic mean are three frequently used operators which are special cases of an ordered weighted averaging operator. The resulting correlation coefficient of the maximum (α equal to zero) and minimum (α equal to infinity) operator is 0.630 respectively 0.461. Rather than taking into account only one of the six indicators, the arithmetic mean operator computes the index scores as the average of all indicator values. In case of linear aggregation (and assuming an equal importance of the indicators) the correlation coefficient becomes 0.735. Moreover, geometric aggregation results in a correlation coefficient of 0.680. From this, it can be concluded that incorporating all indicators in the index causes a better approximation of the final road safety outcome ranking. Furthermore, in case of linear aggregation the degree of agreement appears to be higher than in case of geometric aggregation. However, as stated in Nardo et al. (2005b) the use of linear aggregation requires the indicators to be so-called mutually preferentially independent. This implies that in case there are some synergies or conflicts between the indicators, the index constructed by means of linear aggregation will be biased and not entirely reflect the indicator information. Moreover, a bad performance with respect to one indicator can be compensated by sufficiently high values of other indicators.

Table 27: Spearman's correlation coefficients between index rankings based on a specific aggregation operator and the final outcome ranking

	CORRELATION COEFFICIENT
MAXIMUM	0.630
MINIMUM	0.461
ARITHMETIC MEAN	0.735
GEOMETRIC MEAN	0.680
OWA with α=2	0.714

Apart from these special types of OWA operators, numerous weighting vectors are possible resulting in different index scores and countries' rankings. By means of linguistic formulations restrictions could be obtained for α. The weighting vector will then respect the attitude of experts in terms of aggregation. Based on two linguistic guidelines from a panel discussion of road safety experts, the interval [2;2.65] has been derived to provide an acceptable α. For this interval, the correlation coefficients of the ranking based on the computed index scores and the final outcome ranking lie between 0.69 and 0.71. More specifically, the index ranking resulting from α equal to 2 obtains the highest correlation coefficient, i.e., 0.714. Although the degree of agreement (based on correlation) is still lower than the 0.735 resulting from the arithmetic mean operator, the way of aggregating the road safety performance indicators is considered to be acceptable in the former case as it reflects the attitude of experts. For α equal to 2, the ordered indicator values (starting with the best performance) receive the following weights: 0.03; 0.08; 0.14: 0.19; 0.25; 0.31. This shows that all indicator values are used to some extent (compared to the max and min operator), compensation is only limitedly allowed (contrary to the arithmetic mean operator) and the three worst performances of each country account for 75% of its overall index score. Finally, a corresponding orness value of 0.31 implies rather intolerant behaviour.

8.6 Concluding remarks

In this chapter, the process of aggregating indicators into an index has been discussed. Aggregation is an extensive research domain with some valuable aspects for the road safety index case. In general, aggregation functions or operators can be divided into four broad classes, i.e., averaging functions, conjunctive functions, disjunctive functions and mixed functions. Here, the class of averaging operators in which the index score is bounded by the lowest and highest indicator value has been studied in more detail. More specifically, ordered weighted averaging operators assigning weights to magnitudes of performances appear to be useful aggregation operators as good and bad performances can have a different contribution to the index score. Special types of OWA operators have been discussed and different weighting vectors interpreted.

The aggregation research showed that considering the maximum or minimum operator – and restricting the road safety index to consist of one indicator only – is not justifiable based on numeric (correlation) results; even more importantly is the belief that best indicators have been selected, being relevant and essential parts of the road safety performance index. The degree of complexity is another issue to raise. Other studies dealing with the construction of a composite index often select the most simple methods (i.e., equal weighting and the arithmetic mean operator). Although linear aggregation is very straightforward, one should keep in mind that this way of aggregation is only plausible if no synergies or conflicts exist between the indicators.

Different OWA weighting vectors have been discussed. By changing the α parameter, the impact on the share of weight assigned to good or bad performances was shown. A higher α entails less optimistic and more intolerant behaviour as the bad performances are considered to have a larger impact and are therefore emphasized. The attitude of decision makers or experts can easily be incorporated in the aggregation process thereby enhancing the value and acceptability of the constructed index. The final weighting vector is the one that on the one hand appropriately reflects the aggregation idea of decision makers and/or experts and on the other hand results in a high correlation with the final road safety outcome ranking.

This chapter demonstrated the added value of ordered weighted averaging operators in the creation of a road safety performance index. The next chapter deals with robustness testing. The impact of methodological choices made during the index construction process (with respect to indicator selection, normalisation, weighting and aggregation) on the final countries' ranking will be assessed.

Chapter 9 Testing the robustness of the index[35]

Since the rank of a country can be largely influenced by the decisions taken at the different stages of the index process (i.e., the selection of indicators, the normalisation of the indicator values, the indicator weights and the way of aggregating) the robustness of the index needs to be assessed. In other words, the stability in the output given small changes in the input is tested. Section 9.1 describes the idea of robustness testing in detail. Section 9.2 discusses the theoretical considerations of uncertainty analysis (UA) and sensitivity analysis (SA). The results are discussed in Section 9.3 and Section 9.4 gives concluding remarks regarding the robustness of the road safety performance index.

9.1 Introduction

The robustness of the road safety performance index will be assessed by means of uncertainty and sensitivity analyses. Despite the fact that a road safety performance index can be created based on the information in the previous chapters, the current step is an important one in the development of a sound composite index. The impact on the end result (i.e., the countries' ranking) of choices made during the process offers valuable information. As stated in Nardo et al. (2005a) the iterative use of uncertainty and sensitivity analysis contributes to the well-structuring of the composite index, provides information concerning the robustness of the countries' ranking and identifies ways to reduce the uncertainty in the ranking for a better monitoring and policy.

Based on a review of 24 indexes (in fields such as environment, economy, technology and health) Saisana and Tarantola (2002) conclude that in very few studies the consideration of the uncertainty associated to the methodological approach in creating composite indexes is cited. On the contrary, in Saisana et al. (2005) a comprehensive description of uncertainty analysis and sensitivity analysis is given and the theory is

[35] Related research has been published in: Hermans, E., Van den Bossche, F. and Wets, G. (2009b). Uncertainty assessment of the road safety index. Reliability Engineering and System Safety, 94 pp. 1220-1228.

applied to the technology achievement index. In the knowledge economy indicators project (Nardo et al., 2005a) numerous aggregation systems, weighting schemes, indicator combinations and normalisation techniques are included in the model. The most influencing factors are identified by means of a sensitivity analysis. For the internal market index (Tarantola et al., 2002) the effect of different weights has been studied. A sample of 5,000 random points was generated from the frequency distribution of the twelve indicator weights (based on expert opinions). The sensitivity analysis highlighted the three indicators for which more consensus on their relative merit would result in a lower level of uncertainty. Finally, in the e-business readiness composite indicator developed by Tarantola et al. (2006) the influence of variability in the weights and uncertainty in the imputed data has been studied.

In general, the results of the uncertainty and sensitivity analysis indicate how robust the output is and which of the input factors causes most uncertainty. Similar to other studies (e.g., Nardo et al., 2005a) the output of interest in this case is the average change in the countries' ranking over all possible scenarios with respect to a reference ranking. The first reference ranking – the default index (DI) ranking – is based on index scores computed by means of default methods (such as equal weighting and linear aggregation). Yet, due to the specific nature of the road safety performance index a second reference ranking can be set up, i.e., a final outcome ranking based on the number of road fatalities per million inhabitants. In addition, an analysis on the country level will be performed to show which countries have the most unstable rank under the different methodological options. The input factors of which the effect will be studied relate to the set of indicators to combine, the normalisation technique, the weighting method and the way of aggregating.

Policymakers and other users are interested in the end result rather than the preceding information processing. Nevertheless, given its implications (e.g., taking particular road safety action; following some country's example; etc) the variability in the result due to the methodological decisions should be expressed. In Figure 23, the median rank and the corresponding 5th and 95th percentiles of the distribution of road safety performance index ranks are shown for the 21 countries.

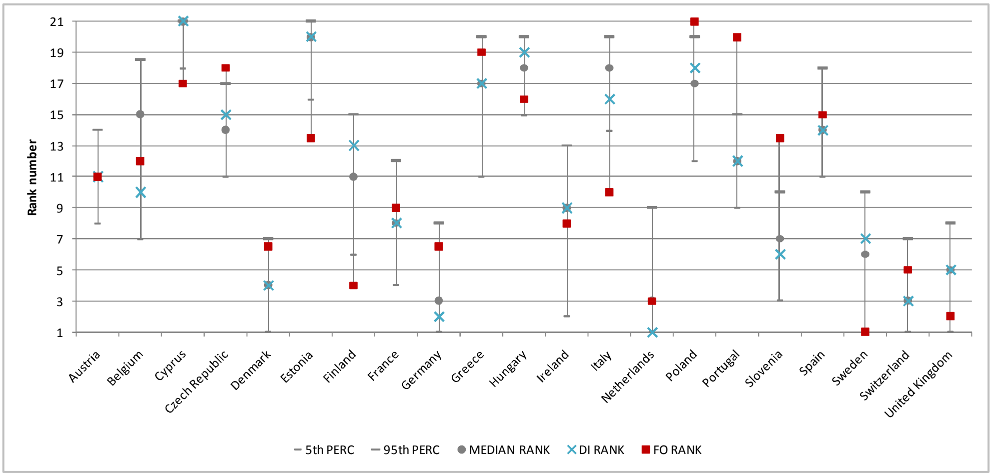


Figure 23: Reference ranks and variability in index ranks per country

The three measures (in grey) – the 5th, 50th and 95th percentile – give an indication of the variability in the possible index ranks for each country due to different sets of indicators and various normalisation, weighting and aggregation methods. The ranks of Belgium, Ireland, Finland and Sweden exhibit a large discrepancy between the different index methodologies. These countries are characterized by excellent as well as poor performances (see Table 10). Consequently, a change in index methodology (e.g., another set of indicator weights) affects these countries more. In addition to the index ranks, the two reference ranks are indicated on the graph for each country. The crosses (in blue) point out the ranks in case the default index methodology is applied. In that case, the Netherlands rank first and Cyprus last. The squares (in red) specify the ranks based on the number of road fatalities per million inhabitants (i.e., Sweden ranking first and Poland last). In the next sections, the average shift in rank over the different index methodologies with respect to the reference rankings will be quantified and the factors causing most uncertainty will be identified. That way, the impact of a change in index methodology on the average shift in rank becomes clear as well as the most influential factors for which the best option needs to be chosen in order to obtain a robust index.

9.2 Method

In general, uncertainty analysis estimates the uncertainty in the output taking into account the uncertainty affecting the input factors. Rather than being a unique value the estimated output represents a distribution of values and elementary statistics – such as mean, standard deviation and percentiles – are used to describe its features (Saltelli et al., 2008). At the same time, sensitivity analysis is defined as the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. Saltelli et al. (2008) present a step-by-step plan for performing a sensitivity analysis on a model:

- *The goal of the analysis and consequently the form of the output function need to be defined. The output of interest is a single quantity whose value is the top-most information that the model is supposed to provide.*

  The rank of each country resulting from a particular set of computed index scores is compared to the reference rank[36] of each country. The output of interest is the average of the 21 absolute differences. It captures the absolute average shift in rank of the entire set of countries relative to the reference ranking. A higher shift indicates a larger uncertainty.

- *A decision has to be made on which input factors to include in the sensitivity analysis.*

  Based on the previous chapters, five input factors are selected to be studied. By making use of triggers one of the possible options for each input factor will be selected each time. The first input factor determines the set of indicators to combine (i.e., indexes consisting of all six indicators and only five indicators are created[37]). The second input factor relates to the normalisation of the data (standardised values, rescaled values and rank numbers

[36] Here, we consider two reference rankings; one with respect to the default index methodology and one with respect to the final outcome indicator.

[37] There are seven options with respect to the first input factor: considering all six indicators; five indicators, i.e., no alcohol and drugs indicator; five indicators, i.e., no speed indicator; five indicators, i.e., no protective systems indicator; five indicators, i.e., no vehicle indicator; five indicators, i.e., no roads indicator; and five indicators, i.e., no trauma management indicator.

are the options). The third input factor determines the method for obtaining indicator weights (i.e., equal weighting, factor analysis, budget allocation and analytic hierarchy process)[38,39]. The fourth input factor decides on the expert in case the previous input factor selected the budget allocation or analytic hierarchy process method[40]. Finally, the fifth input factor decides on the way of aggregating (i.e., linearly; geometrically or using an ordered weighted averaging operator, with α equal to two).

- *For each input factor, a distribution function has to be chosen.*

  All input factors are uniformly distributed between 1 and $r+1$ with r the number of options related to a particular input factor. In other words, [1,8] for the first input factor (indicator selection), [1,4] for the second (normalisation), [1,5] for the third (weighting method), [1,10] for the fourth (expert selection) and [1,4] for the fifth input factor (way of aggregating).

- *A sensitivity analysis method must be chosen.*

  A wide range of methods for sensitivity analysis can be found in literature. First, local and global methods can be distinguished. As local methods explore only one point of the factor's space and factors are changed one at a time, a global analysis is more appropriate in this case. Global methods explore the entire interval of each factor and the effect for a factor is the average over the possible values of the other factors. In recent years,

[38] Data envelopment analysis being the most different weighting method of the five, i.e., an optimization method resulting in weights that do not sum up to one (see Chapter 6) will not be considered as a weighting option here. However, in the results section (9.3.3) some attention is paid to DEA as well.

[39] Given the first input factor, in case of five indicators the weights of the remaining indicators are proportionally adjusted to obtain a set of five weights summing up to one as well.

[40] In Chapter 6, the average BA weights over 11 experts and the average AHP weights over 9 consistent experts were used to represent the importance of the indicators according to a group of experts. In the current chapter, the robustness of the index and the resulting countries' ranking is studied with respect to the different expert opinions. Consequently, all experts providing valuable BA ánd AHP information are included (irrespective of their degree of consistency). To be able to assess the impact of this input factor, 9 experts are included. Together, input factor three and four determine each time which of the 20 possible sets of weights (i.e. EW; FA; BA1; ...; BA9; AHP1; ...; AHP9) is used to construct the index.

global quantitative sensitivity analysis techniques have received considerable attention in literature (Saltelli et al., 2008).

One important and promising class of sensitivity methods is referred to as variance based techniques. Their characteristic of being model-free (i.e., applicable to non-linear and non-additive models) makes them very useful for composite indexes, in which several layers of uncertainty are simultaneously present (Nardo et al., 2005b). Variance based techniques have a number of advantages (Saltelli et al., 2008): they can explore the whole range of variation of each input factor instead of sampling the factors over a restricted number of values, they are able to capture interaction effects apart from the fractional contribution of input factor X_i to the variance of the model output Y, are easy to interpret, are quantitative, etc. Therefore, a variance based sensitivity analysis method is used here.

- *The input sample is generated.*

  A $N \times 5$ input sample is produced. Given the use of triggers and the uniform distribution of the five input factors, the number of methodological scenarios (N) determines the number of rows in the input sample.

- *The model needs to be elaborated on the generated sample and the output produced.*

  The five values in each row of the input sample have to be converted into one output score. First, the values need to be translated to make clear which indicators, which normalisation technique, which weighting method, which expert (if needed) and which way of aggregation is to be used in the construction of the index. The index value for all countries in the data set results from the aggregation of the normalised set of selected indicators and the corresponding weights. Based on their index score, each country is assigned a rank number between 1 and 21. The output value is the average of the 21 absolute differences between the rank based on the computed index scores and the reference rank. N output values are obtained with respect to a particular reference ranking. The first reference ranking – referred to as the default index ranking – is composed using the most common and/or simple option for each input factor. In particular, all six indicators are taken into account, standardised indicator values are used, equal weighting is applied and the aggregation occurs in a linear way. Countries' rankings based on index scores computed in another way are compared to the default index ranking in order to assess the impact of the index methodology. Secondly, the similarity between the possible index rankings and the final outcome ranking is checked. As opposed to indexes developed in other fields, the common ranking based on the number of road fatalities per million inhabitants is interesting to use as a reference ranking. For each methodological scenario, the resulting index ranking will be compared to the ranking based on the number of road fatalities per million inhabitants.

- *Finally, the model output is analysed and conclusions are drawn.*

  The distribution of the output informs us about the variability in rank of the index methodologies with respect to the reference ranking. In addition, some sensitivity indexes are computed. Saltelli et al. (2008) provide detailed information on sensitivity indexes. S_i represents the first-order effect of input factor X_i on output Y. The total effect term S_{Ti} gives information on the total effect of input factor X_i on output Y. The characteristics of the first-order and total effect sensitivity index are shown in Table 28 (Saltelli et al., 2008). Based on these sensitivity indexes insight into the most influential and possibly non-influential input factor(s) can be gained.

Table 28: Summary information on the first-order and total effect sensitivity index

	DESCRIPTION
$V(Y)$	The unconditional variance quantifying the uncertainty of the model output Y
$V(E(Y \mid X_i)$	The expected reduction in the output variance if X_i could be fixed to one of its values $(x_i^* = 1,\ldots,r)$[41]
$S_i = \frac{V(E(Y \mid X_i)}{V(Y)}$	The first-order sensitivity index indicates the relative importance of an individual input factor X_i in driving the uncertainty
$S_i = \text{high}$	X_i is an influencing input factor that deserves careful consideration in order to reduce the output uncertainty
$E(V(Y \mid X_{-i})$	The expected amount of output variance that would remain unexplained if all input factors but X_i (i.e., X_{-i}) could be fixed to one of their values[42]
$S_{Ti} = \frac{E(V(Y \mid X_{-i})}{V(Y)}$	The total effect sensitivity index captures the total effect of X_i on Y
$S_{Ti} - S_i$	This measure indicates how much X_i is involved in interactions with other input factors
$S_{Ti} = 0$	X_i is non-influential, i.e., not important neither singularly nor in combination with other input factors
$\sum_i S_{Ti} > 1$ or equivalent $\sum_i S_i < 1$	Interactions exist between the input factors

Saltelli et al. (2008) state that by computing all S_i and S_{Ti} terms, a fairly complete and parsimonious description of the model can be obtained in terms of its global sensitivity properties. The uncertainty and sensitivity results are discussed in Section 9.3.

[41] Recall that r is the number of options related to a particular input factor. $x_1^* = 1,2,3,4,5,6,7$; $x_2^* = 1,2,3$; $x_3^* = 1,2,3,4$; $x_4^* = 1,2,3,4,5,6,7,8,9$; $x_5^* = 1,2,3$.

[42] For i=1, each combination of possible values of the 2nd, 3rd, 4th and 5th input factor (e.g., F_2=1,F_3=1,F_4=1,F_5=1 or F_2=3,F_3=4,F_4=9,F_5=3) is considered (regardless of the value of the first input factor).

9.3 Results

Following the step-by-step plan discussed in the previous section, an uncertainty and sensitivity analysis is performed for the road safety performance index. A uniform probability distribution function is assigned to the indicator set (six or five road safety performance indicators), the normalisation technique (standardised, rescaled or rank data), the weighting method (equal weighting, factor analysis, budget allocation or analytic hierarchy process), the expert selection (expert 1 to 9) and the way of aggregating (linearly, geometrically or based on an ordered weighted averaging operator with α equal to two). Next, values are drawn from these distributions. Each possible scenario appears once in the input sample M consisting of N rows and F (=5) columns. However, it should be noted that not all 2268 (i.e., 7x3x4x9x3) combinations of input triggers are incorporated in the input sample. In particular, the combination of geometric aggregation (i.e. $F_5 = 2$) and standardised scores ($F_2 = 1$) – occurring 252 times – would result in invalid results (due to the fact that negative values are raised to a particular power, e.g., $-2^{0.3}$).

As illustrated in Figure 24 for the 700th row of sample M with respect to the first (default index) reference ranking the set of triggers in each row produces 21 road safety index scores. Based on these scores a road safety index rank is assigned to each country (a higher index score implying a better rank). Using the reference ranking of the 21 countries, the absolute difference between the two rank numbers is calculated for each country. Subsequently, these 21 values are averaged resulting in the output value corresponding to that specific sample row. This process is repeated for the next row in the sample. In the end, N output values are obtained. The results of comparing the index rankings against the default index ranking (the first reference ranking) are discussed in Section 9.3.1. In this case, N equals 2007 (i.e., 2268-252-9) output values; the default methodology (i.e., F1=1; F2=1; F3=1; F5=1; occurring 9 times) is not considered as it is the reference ranking[43]. In Section 9.3.2 the results with respect to the second reference ranking (the final

[43] Otherwise, this combination of input triggers would produce an output value of zero, thereby favourably influencing the average shift in rank over all possible scenarios.

outcome ranking) are given. In that case, 2016 (i.e., 2268-252) output values are evaluated.

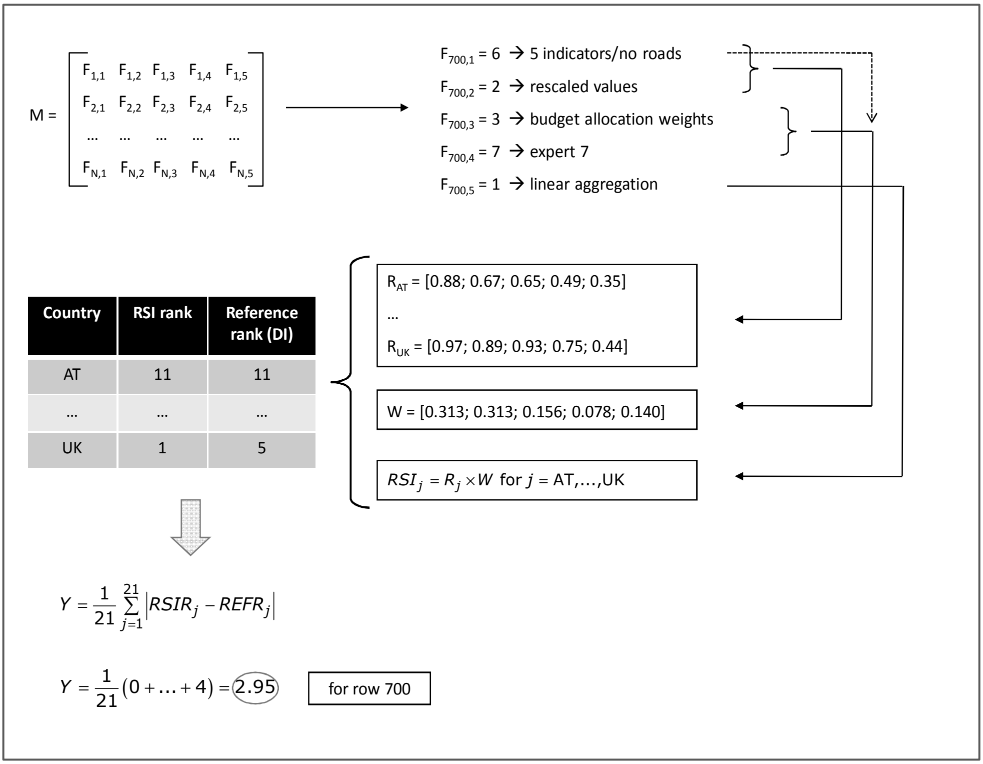
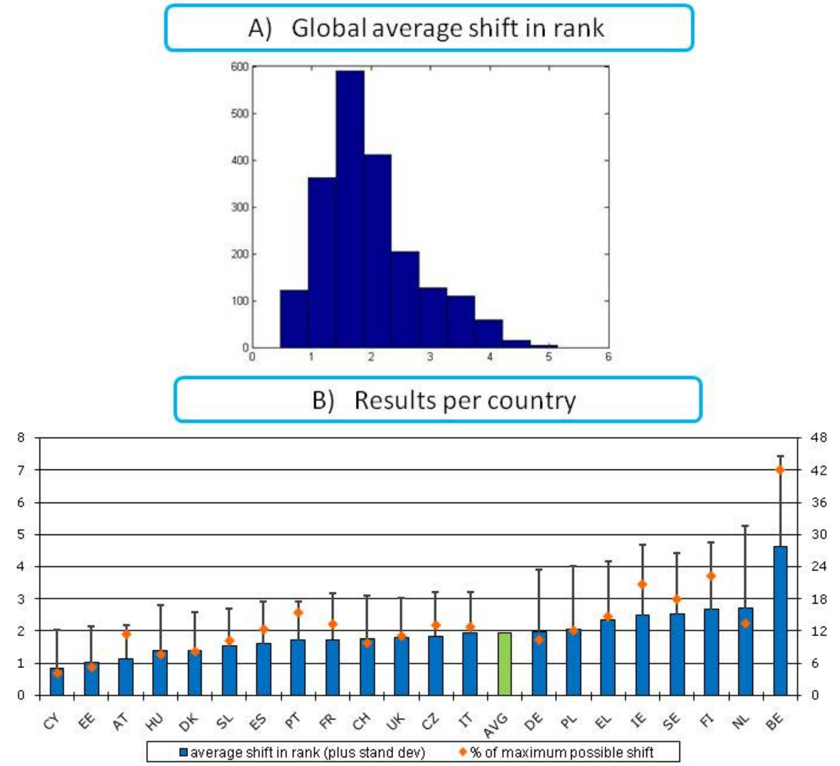


Figure 24: Computing the output of interest

9.3.1 The default index ranking

The 2007 obtained output values are used to draw conclusions regarding the uncertainty and sensitivity of the countries' ranking with respect to the default index ranking (based on the linear aggregation of the six standardised indicator values and equal weights). The output distribution is shown by the histogram in Figure 25 (A). The absolute average shift in rank over all 21 countries has an average of 1.96 positions with a variance of 0.68, a minimum of 0.48 and a maximum of 5.14. The index ranking resemblances the reference ranking best in case the six rescaled values are equally weighted and linearly aggregated.



C) Sensitivity indexes

Input factor	First-order sensitivity index	Total effect sensitivity index	Difference between the two indexes
Indicator selection	0.083	0.239	0.156
Normalisation technique	0.017	0.190	0.173
Weighting method	**0.349**	**0.644**	**0.295**
Expert selection	0.058	0.296	0.238
Way of aggregating	0.074	0.280	0.206
Sum	**0.581**	**1.649**	**1.068**

Figure 25: Uncertainty and sensitivity results regarding the default index ranking

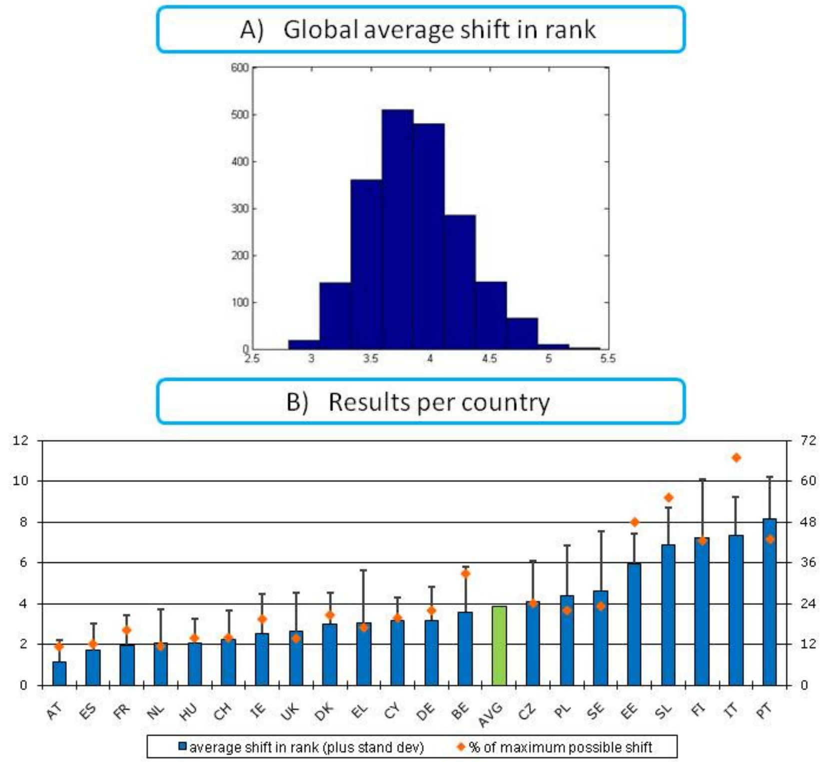
Moreover, the results per country are shown in Figure 25 (B). The average shift in rank (and the standard deviation) shows that Belgium is the country for which an index methodology different from the default methodology results in a very dissimilar rank. Figure 23 already revealed the discrepancy between the median rank over all methodological scenarios and the default index rank of Belgium. A detailed analysis shows that the high average shift in rank is mainly due to a change in the weighting method in general and the use of BA or AHP weights in particular. On average, these participatory methods stress the alcohol and drugs domain and the speed domain whereas the vehicle and trauma management values limitedly contribute to the index score. This particularly affects Belgium scoring bad on the behavioural dimensions and well on the vehicle, roads and trauma management dimensions. In addition, the average shift in rank of a country is studied in relative terms. Countries with a reference rank of 1 or 21 have a maximum possible shift of 20 whereas this is only 10 for countries with a reference rank of 11. The relative average shift in rank expressed as share of the maximum possible shift in shown in Figure 25 (B) as well. The 42% of Belgium again illustrates the high impact of a change in index methodology on its shift in rank. On the contrary, the various methodological scenarios cause a negligible variability in average shift for Cyprus and Estonia. In other words, the relative position of these countries is hardly affected by the methodology applied for creating a road safety performance index.

Finally, using the formulas in Table 28, five first-order and total effect sensitivity indexes – S_i respectively S_{Ti} – are calculated with respect to the global average shift in rank. S_i captures the fractional contribution to the model output variance due to the uncertainty in X_i and S_{Ti} concentrates all the interactions involving factor X_i in one single term. As shown by the tabulated values in Figure 25 (C) the weighting method is the most influencing factor. In addition, the last column shows that the input factor 'expert selection' is also involved in interactions with the other input factors to some extent. The tested normalisation options seem to have the smallest effect. Nevertheless, in general it can be concluded that all five factors appear to have an influence on the variability in countries' ranking.

9.3.2 The final outcome ranking

The 2016 output values are used to obtain the uncertainty and sensitivity outcomes with respect to the ranking based on the number of road fatalities per million inhabitants. The output distribution is visualised by the histogram in Figure 26 (A). The absolute average shift in rank over all 21 countries has an average of 3.87 positions with a variance of 0.16, a minimum of 2.81 and a maximum of 5.43. A detailed look at the results demonstrates that the largest absolute average shift in rank occurs in case the composite index consists of five indicators (no protective systems indicator), rescaled data, budget allocation weights from expert 8 (assigning one third of the weight to the roads indicator) and geometric aggregation. Figure 26 (B) shows the disaggregation of the global absolute average shift in rank. For each country, the average over the 2016 values of the absolute difference between their rank based on the road safety index scores and their final outcome rank is shown in ascending order. Some countries exhibit a rather high average shift and/or standard deviation. For example, a shift of 14 positions is encountered for Portugal in case five indicators (no speed indicator) are combined using ordinal data, the budget allocation weights from expert 8 and geometric aggregation. Based on Figure 26 (B) it can be concluded that five countries – Portugal, Italy, Finland, Slovenia and Estonia – have a high average shift in rank over the different methodological options with respect to their final outcome rank. Figure 23 also showed that the distribution of the index ranks over the various methodological scenarios does not match the final outcome rank for these countries. At the same time, the different methodological scenarios studied here only slightly affect the difference between the index rank and the final outcome rank for Austria.

The input factors causing most uncertainty can be identified by means of the first-order and total effect sensitivity indexes shown in Figure 26 (C). It is clear that the indicator selection is the most influencing input factor; it has a first-order sensitivity index that is on average five times higher than the one of the other four input factors. Moreover, the last column shows that all input factors are affected by interactions. Based on the total effect sensitivity indexes it can be stated that the selection of indicators to include in the performance index as well as the weighting method applied cause most variability in the output.



C) Sensitivity indexes

Input factor	First-order sensitivity index	Total effect sensitivity index	Difference between the two indexes
Indicator selection	**0.218**	**0.533**	0.315
Normalisation technique	0.097	0.335	0.238
Weighting method	0.026	**0.398**	**0.372**
Expert selection	0.054	0.342	0.288
Way of aggregating	0.019	0.311	0.292
Sum	**0.414**	**1.919**	**1.505**

Figure 26: Uncertainty and sensitivity results regarding the final outcome ranking

9.3.3 Discussion

During the index construction process, several methodological choices and assumptions need to be made. The uncertainty and sensitivity of the end result related to decisions concerning indicator selection, normalisation technique, weighting method, expert selection and aggregation have been assessed. For each scenario, the countries' ranking based on the road safety performance index scores was compared to a reference ranking. For the road safety index case, two reference rankings were considered. The first reference ranking was used to assess the influence of a change in index methodology compared to the default (i.e., the most common/simple) methodology and the second reference ranking to test the accordance between the road safety performance index ranking and the final outcome ranking.

The previous sections showed different uncertainty and sensitivity results with respect to the reference ranking. The absolute average shift in rank over all 21 countries compared to the default index ranking has an average of 1.96 positions with a variance of 0.68 whereas compared to the final outcome ranking the average equals 3.87 positions and the variance 0.16. This absolute average shift with respect to the second reference ranking indicates that there still appears to be some discrepancy with the final outcome ranking. Most likely, we need to incorporate more road safety indicators, some indicators may only be relevant for a certain group of countries or the quality of the data may be rather poor.

The disaggregation of the global average shift in rank per country revealed that the robustness of the rank number of a country with respect to a change in index methodology greatly differs between the 21 countries. Considering both reference rankings, the various methodological scenarios caused an average shift in rank of more than 4 positions for one third of the countries.

Apart from obtaining insight into the degree of uncertainty, first-order and total effect sensitivity indexes were computed to assess which input factor(s) accounts for most uncertainty in the output. In both cases, the first-order sensitivity index clearly indicated the most influential input factor. Regarding the default index ranking, the weighting method ($S_3 = 0.349$) seemed to cause most uncertainty; regarding the final outcome ranking, the uncertainty could be reduced most if more effort was directed in the selection of indicators to combine ($S_1 = 0.218$). The sum of the five first-order sensitivity indexes (0.581 respectively 0.414) is smaller than one implying the existence of interactions between the input factors. Therefore, the calculation of total effect sensitivity indexes is justified. The relative differences between the total effect and first-order sensitivity indexes indicate that all input factors are involved in interactions to some extent. The sensitivity results in other index studies (e.g., Nardo et al., 2005a) showed that the indicator selection, the weighting method and the choice of the expert are often important factors. Based on the results it can be concluded that more consensus on the weighting method would result in a more robust index ranking. In addition, the discrepancy with the final outcome ranking could be mainly reduced by focusing on the selection of indicators.

Various scenarios for creating a road safety performance index have been evaluated in this chapter: including all six best available indicators or only five; using standardised, rescaled or rank data; applying equal weights, factor analysis weights, budget allocation weights or analytic hierarchy process weights; considering the BA or AHP weights from one of the nine experts; and aggregating in a linear, geometric or ordered weighted averaging way. The data envelopment analysis method was not included and will be briefly discussed now. In particular, the absolute average shift in rank with respect to both reference rankings is computed for a number of DEA scenarios. The optimization model described in Chapter 6 (Eq. 6–2) will be applied. The road safety performance index scores follow from the linear aggregation of the raw indicator values and the most favourable weights respecting all restrictions. Therefore, DEA as weighting method (input factor 3) renders the normalisation technique (input factor 2), the expert choice (input factor 4) and the way of aggregating (input factor 5) redundant. Considering the only relevant input factor (the indicator selection; input factor 1) and its seven options gives the following results: the absolute average shift in rank with respect to the default index ranking has an average of 2.63 positions whereas compared to the final outcome ranking the average equals 3.91 positions. Looking again at the average values over all tested scenarios (1.96 respectively 3.87) it can be seen that the result with respect to the final outcome reference ranking is quite similar while on average the seven DEA scenarios differ more from the default index methodology than the other scenarios did. Moreover, it can be concluded that the use of all six indicators in the data envelopment analysis model results in the lowest average shift in rank with respect to both reference rankings.

In general, the various methodological scenarios generate a quite stable ranking of countries. In the selection of the final methodology (see

Section 10.1) the qualitative appraisal of the different options should be paid attention to. Methods with interesting advantages for the road safety case and offering additional insight in the mechanisms underlying the end result are to be opted for.

9.4 Concluding remarks

The robustness of the road safety performance index has been studied in this chapter. Given the multistage index process, the developer is left with a number of methodological decisions that might have large implications for the final countries' ranking. In other words, a country can have very different ranks depending on the index methodology applied (more specifically, the selected set of indicators, normalisation technique, weighting method, expert selection and way of aggregating). Given the impact of an index ranking (e.g., in terms of communication and benchmarking) the robustness should be assessed and evaluated. By means of uncertainty and sensitivity analyses the impact of the methodological options can be quantified and the factors causing most uncertainty identified. Here, the influence of five aspects on the absolute average shift in countries' ranking with respect to a reference ranking has been assessed.

On average, the tested methodological scenarios resulted in an absolute average shift in rank of 1.96 positions compared to the default index ranking. Moreover, the ranking relating to all methodological scenarios differed the most from the ranking based on the linear aggregation of six equally weighted standardised indicator values due to a change in weighting method. Therefore, the selection of the weighting method is crucial and needs careful evaluation and justification. A second reference ranking was used in order to assess the degree of agreement with the final outcome ranking. An absolute average shift in rank over all methodological scenarios of 3.87 positions was found. The selection of indicators appeared to be the most influencing factor. Consequently, the decision of which indicators to combine in the road safety performance index is essential for obtaining robust results.

This chapter illustrated the use of uncertainty and sensitivity analyses. In the next chapter, the final index methodology is described. Based on the best available indicator set, final road safety performance index scores will be computed and discussed. Given the robustness results, the selection of the weighting method will receive special attention.

Chapter 10 Computing, evaluating and visualising final index scores

At this point, we have reached the last step of the index construction process: the computation, evaluation and visualisation of the final index results. Following the information acquired in the previous steps and in particular the qualitative and quantitative assessment of the different methodological options, the final index methodology for the creation of a composite road safety performance index is described in Section 10.1. Using this methodology, the indicator values of each of the 21 European countries will be combined in a final road safety performance index score (Section 10.2). Next, these scores are evaluated – i.e., compared to the results of other, related studies of indexes or indicators – and visualised (Section 10.3). Concluding remarks are given in Section 10.4.

10.1 Final index methodology

To obtain the final road safety performance index scores a decision concerning indicator selection, normalisation, weighting and aggregation is required. In the previous chapters, these steps were handled. In particular, in Chapters 3, 4 and 5 possible indicators were evaluated, best available indicator values gathered and described by means of univariate and multivariate analyses. The topic of weighting was examined in Chapter 6 where five common methods – factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting – were studied in detail. Various aggregation operators were described in Chapter 8 with special attention given to the class of ordered weighted averaging operators. Within each step, various options interesting for the road safety index case exist, each having some benefits and limitations. The effect of the chosen index methodology on the countries' ranking has been studied by means of uncertainty and sensitivity analyses (Chapter 9).

The final methodology for the creation of the road safety performance index is presented here. Within the different steps of the index process the 'best' options are selected. In this respect, a quantitative criterion such as the degree of correlation between the index ranking and the final outcome ranking can be employed (see Section 2.9). However, the dependency of these results on the data used should be kept in mind. Primarily, a valuable and acceptable index should be created. Therefore, a more qualitatively oriented approach in which methods offering additional insights are preferred, will be followed to select the final index methodology.

A road safety performance index will be created taking the six best available indicators into account. More specifically, the index score will represent the overall performance of each of the 21 countries concerning alcohol and drugs (the percentage of surveyed car drivers respecting the blood alcohol content limit), speed (the percentage of surveyed car drivers with a driving speed below the speed limit in built-up areas), protective systems (the seat belt wearing rate in front of cars and vans), vehicle (the percentage of cars that are maximum five years old), roads (the density of motorways) and trauma management (the total health expenditure as share of the gross domestic product). Moreover, for the current data set the robustness analysis showed the influence of the method used for determining the indicator weights. Therefore, in the development of the final index methodology, a decision upon the weighting method is made first. In Chapter 6, five weighting methods have been studied in detail. Although each method has a number of advantages, disadvantages and requirements (see Table 15) in the current context the data envelopment analysis method is preferable to the other four methods because of the following reasons:

- It is an objective method deriving indicator weights from the data
- Prior knowledge or accepted views (e.g., from road safety experts) can be easily included as restrictions in the model to guarantee realistic and acceptable weights
- It is an optimization method in which the indicator weights result in the most optimal index score
- It can handle raw values rendering the normalisation technique redundant and eliminating its effect on the final index scores and the countries' ranking
- The model output offers additional information, e.g., benchmark countries, good and bad aspects, etc

- It can take indicator information over time into account to assess the degree of progress as well as the change regarding the benchmark
- Country-specific models as well as an overall model can be formulated

Hence, compared to the other weighting methods, DEA holds a lot of new perspectives. The next aspect of the index methodology relates to the way of aggregating. Chapter 8 demonstrated how the aggregation idea of decision makers and/or experts could be reflected in the index scores. Based on linguistic formulations (Section 8.4), an ordered weighted averaging operator with an α parameter lying between 2 and 2.65 has been determined implying that the worst performances of each country are emphasized. This information can be incorporated in the DEA model by adding an orness[44] restriction. α equal to 2 respectively 2.65 corresponds to an orness of 0.306 and 0.236. However, adding a restriction that the orness should lie between 0.236 and 0.306 has two major implications. First, ordered data values are to be used because the orness is defined in terms of decreasing performance. The first weight $(\overrightarrow{w_1})$ no longer corresponds to the alcohol and drugs performance but to the best performance over the six indicators. In order to deduce the sequence of performance for a particular country (e.g., speed > vehicle > alcohol and drugs > roads > trauma management > protective systems) normalised data (instead of raw values) are required. In Chapter 5, three techniques have been discussed. In case the raw values are transformed into rank numbers some valuable information is lost (i.e., the size of the difference in value between two following rank numbers). Here, the raw values will be rescaled rather than standardised so we dispose of positive values lying between zero and one and obtain index scores within an interpretable interval. Second, the orness is computed for a weighting vector summing up to one. Therefore, the sum of the six ordered weights is forced to equal one. Given the fact that all indicator values are normalised between zero and one (first implication) and that

---

[44] $Orness = \frac{1}{I-1}\sum_{i=1}^{I}(I-i)\overrightarrow{w_i} = \frac{1}{5}\times(5\times\overrightarrow{w_1}+4\times\overrightarrow{w_2}+3\times\overrightarrow{w_3}+2\times\overrightarrow{w_4}+1\times\overrightarrow{w_5})$

the six ordered weights sum up to one (second implication) all composite index scores will lie within the [0,1] interval.

In contrast to the DEA models discussed in Chapters 6 and 7, the involvement of an orness restriction leads to the use of normalised, ordered data and ordered weights summing up to one. Nevertheless, the meaning of a specific indicator value – i.e., whether it relates to alcohol/drugs, speed, protective systems, vehicle, roads or trauma management – is still relevant and taken into account by means of pie share restrictions. The upper and lower limits are the same as before[45]; irrespective of the relative position – best, 2nd, 3rd, 4th, 5th or worst performance – the share of the alcohol and drugs indicator in the index should be between 7.7% and 37.1%; for speed between 14.9% and 48.8%; for protective systems between 7.2% and 29.7%; for vehicle between 2.2% and 14.9%; for roads between 1.5% and 27.7%; and for trauma management between 2.2% and 23.9% (see also Table 20). The six weights both correspond to a relative position (where generally worst performances receive a higher weight) and a meaning (taking into account the relative importance assigned to the six indicators by a panel of road safety experts). $\overrightarrow{w_{11}}$ refers to the best performance of Austria, i.e., alcohol and drugs whereas $\overrightarrow{w_{12}}$ (Belgium being the second country in the data set) is linked to roads. Since the ordered weights correspond to different indicators for the countries, country-specific models are formulated. In addition, the weights of the country under study will not be substituted for the other countries as for example the best/alcohol and drugs weight of Austria would be multiplied with the best/roads value of Belgium. We focus on the most optimal index score of the country under study resulting from the best possible weights respecting all (orness as well as pie share related) restrictions.

A final remark involves data values equal to zero. Rescaling transforms the minimum raw value of each indicator into zero. Eleven times the normalised value appears to be 0.00 (Cyprus and Estonia are

[45] The average of the two lowest (respectively highest) weights of eleven budget allocating road safety experts is used as lower (respectively upper) limit for each indicator.

characterized by a share of zero values of 50%). Consequently, the lower pie share limits cannot be satisfied and are not considered in that case (e.g., the share of the speed indicator in the index score of Belgium is zero rather than equal to or larger than 14.9%). However, these worst performances are reflected in the index scores consisting of all six rescaled, ordered values and the corresponding weights. (Eq. 10–1) represents the algebraic model used to compute the final road safety performance index score *(RSPI)* for country j $(j = 1,\ldots,n)$.

$$RSPI_j = \max_{\overrightarrow{w_{ij}}} \sum_{i=1}^{I} \overrightarrow{r_{ij}}\,\overrightarrow{w_{ij}} \qquad \text{(Eq. 10–1)}$$

$$\text{s.t.} \quad \sum_{i=1}^{I} \overrightarrow{w_{ij}} = 1$$

$$0.236 \le \frac{1}{I-1} \sum_{i=1}^{I} (I-i)\overrightarrow{w_{ij}} \le 0.306$$

$$L_m \le \frac{r_{mj} w_{mj}}{\sum_{i=1}^{I} \overrightarrow{r_{ij}}\,\overrightarrow{w_{ij}}} \le U_m$$

$$\overrightarrow{w_{ij}} \ge 0$$

with I = number of indicators

$\rightarrow$ = ordered value

r = rescaled value

w = weight

$m = \{\text{alc;sp;ps;veh;road;tm}\}$

L = lower limit

U = upper limit

As shown in (Eq. 10–1) the road safety performance index score of a country consists of ordered rescaled indicator values (i.e., values between zero and one in decreasing order) and ordered weights (i.e., the first weight corresponding to the best performance). To incorporate the aggregation idea deduced by means of linguistic formulations – i.e., punishing bad performances – an orness value lying between 0.236 and 0.306 was postulated. In addition to this, the expert knowledge regarding the importance of the different indicators was taken into account. More specifically, the share of each of the six indicators in the total index score was restricted by a lower and upper limit. Only one set of weights is used; for a particular country, each weight w_m corresponds to a certain weight $\overrightarrow{w_i}$. The same applies to the indicator values of the country, $\overrightarrow{r_i}$ and r_m (e.g., $\overrightarrow{r_{11}} = r_{alc1}$ for Austria). The subscript m is used to indicate that the one but last restriction relates to the meaning of the indicator instead of its position in the ordering.

To conclude, the road safety performance index methodology presented in this section has been set up taking the knowledge acquired in the previous chapters of this manuscript into account. To the promising data envelopment analysis model an orness restriction has been added. That way, the most optimal index score for a country – lying between zero and one – is obtained respecting the weight restrictions concerning the importance of each indicator and the magnitude of the performance. At the same time it should be noted that this methodology, like any other, has limitations. On the one hand, the indicator weights can be justified as a combination of objective (data) and subjective (expert) information; on the other hand, the reliance on the data set and the expert evaluations can be considered as a limitation of the technique. Here, a best available yet imperfect data set is used: only six risk domains are considered, each risk domain is captured by one indicator and some proxy indicators and self-reported data are to be used at this moment. Furthermore, the use of expert panels and the related accuracy of the findings has been criticised in Washington et al. (2009). Here, expert's opinions are indirectly incorporated in the methodology as upper and lower boundaries in order to direct the most optimal solution based on objective information (data) towards acceptable (realistic) index results. Finally, the aim of the country-specific models is to determine the most optimal index score (under the imposed restrictions) for each country rather than to take into account the performance of the other countries in the data set.

10.2 Computing final index scores

The methodology described in the previous section will be used to compute the 21 final road safety performance index scores. Each index score combines six best available indicator values. The algebraic model presented in (Eq. 10–1) is elaborated for the first country in the data set.

From the six rescaled indicator values of Austria (0.88; 0.67; 0.65; 0.49; 0.32; 0.35) it can be seen that the performance diminishes as follows: alcohol/drugs – speed – protective systems – vehicle – trauma management – roads. In other words, the alcohol and drugs lower and upper limit are applicable to the firstly ordered values $\overrightarrow{r_{11}}$ and $\overrightarrow{w_{11}}$; the speed limits to the secondly ordered values $\overrightarrow{r_{21}}$ and $\overrightarrow{w_{21}}$; ... ; and the roads limits to the lastly ordered values $\overrightarrow{r_{61}}$ and $\overrightarrow{w_{61}}$. Filling out the rescaled, ordered indicator values of Austria and the indicator limits results in (Eq. 10–2).

$$RSPI_{AT} = \max_{\overrightarrow{w_{i1}}} \left(0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}\right) \quad \text{(Eq. 10–2)}$$

$$\begin{aligned}
\text{s.t.} \quad & \overrightarrow{w_{11}}+\overrightarrow{w_{21}}+\overrightarrow{w_{31}}+\overrightarrow{w_{41}}+\overrightarrow{w_{51}}+\overrightarrow{w_{61}} = 1 \\
& 0.236 \le \frac{1}{5}\left(5\overrightarrow{w_{11}}+4\overrightarrow{w_{21}}+3\overrightarrow{w_{31}}+2\overrightarrow{w_{41}}+1\overrightarrow{w_{51}}\right) \le 0.306 \\
\text{[ALC]} \quad & 0.077 \le \frac{0.88\overrightarrow{w_{11}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.371 \\
\text{[SP]} \quad & 0.149 \le \frac{0.67\overrightarrow{w_{21}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.488 \\
\text{[PS]} \quad & 0.072 \le \frac{0.65\overrightarrow{w_{31}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.297 \\
\text{[VEH]} \quad & 0.022 \le \frac{0.49\overrightarrow{w_{41}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.149 \\
\text{[ROAD]} \quad & 0.015 \le \frac{0.32\overrightarrow{w_{61}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.277 \\
\text{[TM]} \quad & 0.022 \le \frac{0.35\overrightarrow{w_{51}}}{0.88\overrightarrow{w_{11}}+0.67\overrightarrow{w_{21}}+0.65\overrightarrow{w_{31}}+0.49\overrightarrow{w_{41}}+0.35\overrightarrow{w_{51}}+0.32\overrightarrow{w_{61}}} \le 0.239 \\
& \overrightarrow{w_{i1}} \ge 0
\end{aligned}$$

(Eq. 10–2) results in an index score of 0.461. The ordered weighting vector (0.040; 0.102; 0.211; 0.039; 0.209; 0.399) has an orness value of 0.306. This weighting vector both represents the rather intolerant aggregation idea of the experts (i.e., a degree of orness in which weaker performances are emphasized) and the importance of the different indicators (i.e., limits based on budget allocating expert results). Moreover, the lower limit with respect to alcohol/drugs and speed is binding as well as the upper limit regarding protective systems and roads.

The results for the other 20 European countries are obtained in a similar way. The road safety performance index scores of the 21 countries are shown in Table 29 in decreasing order. The largest index score (0.692) belongs to the Netherlands. Cyprus and Estonia are assigned an index score of zero. 19 countries have an orness equal to 0.306[46]. Next, the evaluation and visualisation of the index scores are discussed.

Table 29: Countries' ranking based on road safety performance index scores

	Road safety performance index score
Netherlands	0.692
Germany	0.680
Switzerland	0.674
Denmark	0.597
Slovenia	0.596
Unt Kingdom	0.552
France	0.539
Ireland	0.486
Sweden	0.465
Austria	0.461
Belgium	0.393
Portugal	0.378
Finland	0.357
Spain	0.340
Czech Rep	0.319
Greece	0.290
Italy	0.288
Poland	0.256
Hungary	0.209
Cyprus	0.000
Estonia	0.000

10.3 Evaluating and visualising final index scores

In Section 10.1 the reasoning behind the development of the final index methodology has been discussed. Six best available road safety performance indicators have been combined in an index using a data envelopment analysis optimization method with an orness restriction.

[46] Cyprus and Estonia on the other hand have an orness of 0.236.

This section provides information to the end users about the relationship between the constructed index and other, related research and visualises the main results in a comprehensible way.

Based on Table 29, three groups of countries are distinguished (see also Figure 27):

1. Countries with a high road safety performance index score: the Netherlands, Germany, Switzerland, Denmark, Slovenia, United Kingdom and France (displayed in green)
2. Countries with an average road safety performance index score: Ireland, Sweden, Austria, Belgium, Portugal, Finland and Spain (yellow)
3. Countries with a low road safety performance index score: Czech Republic, Greece, Italy, Poland, Hungary, Cyprus and Estonia (red).

Figure 27: Coloured map on road safety performance in Europe

The relationship between the composite road safety performance index developed here and other, related research is determined. First, the link between the intermediate outcome level and the final outcome level is studied. More specifically, the degree of similarity between the composite road safety performance index ranking and the final outcome ranking will be assessed. The final outcome values vary between 59 (Sweden; rank 1) and 149 (Poland; rank 21) road fatalities per million inhabitants with an average of 106 (2003 data). Next, the correlation between the road safety performance index and other indexes (outside yet slightly linked to the road safety field) will be computed. In particular, the well-known human development index (HDI) (United Nations, 2005) and the corruption perceptions index (CPI) (Lambsdorff, 2003) are considered. The human development index measures the average achievements in a country in three basic dimensions of human development: a long and healthy life, knowledge and a decent standard of living. The average HDI value over the 21 countries equals 0.919; Sweden has the highest score (0.949); Estonia the lowest (0.853). The corruption perceptions index ranks countries by their perceived levels of corruption, as determined by expert assessments and opinion surveys. CPI scores range between ten (i.e., highly clean) and zero (i.e., highly corrupt). Finland was assigned the highest score (9.7) and Poland the lowest (3.6). The average CPI score over the 21 countries is 6.9.

The results of comparing the road safety performance index ranking to the final outcome, human development index and corruption perceptions index ranking are discussed below. We start the evaluation with the SUNflowerNext report, i.e., the most recent research on benchmarking road safety across Europe.

In the SUNflowerNext report (Wegman et al., 2008) 27 European countries are grouped based on five trials to produce a combined safety index (using principal components analysis and factor analysis). Moreover, a set of 7 final outcome indicators, 2 background variables, 7 intermediate outcome indicators (of which 4 are related to the vehicle domain) and 5 policy performance indicators is considered. Finally, five groups of countries are identified:

- the group with the highest level of safety performance: Sweden, Norway, France, Great Britain and Germany
- the group with a relatively high level of safety performance: Switzerland, the Netherlands, Finland, Denmark, Ireland, Austria, Luxembourg and Malta

- the group with a medium level of safety performance: Cyprus, Slovenia, Portugal, Belgium and Spain
- the group with a relatively low level of safety performance: Estonia, Slovakia, Greece and Czech Republic
- the group with a low level of safety performance: Latvia, Hungary, Poland, Lithuania and Italy

In case we consider three groups of countries (i.e., high or relatively high; medium; relatively low or low) and focus on the 21 European countries used in our research we can compare the classification of the SUNflowerNext study with the classification based on the road safety performance index (presented in Figure 27). Nevertheless, it should be noted that the number and type of indicators strongly differs between both studies as well as the methodology underlying the grouping of the countries.

It can be concluded that 15 of the 21 countries are classified exactly the same. The opposite is true for Slovenia, Ireland, Sweden, Austria, Finland and Cyprus. In the SUNflowerNext research, the first (or best) class consists of most countries whereas the number of countries per class according to the index classification is seven. In the latter classification, Ireland, Sweden, Austria and Finland belong to the second class with an average road safety performance index score. Moreover, Slovenia – belonging to the first group based on its road safety index score – and Cyprus – belonging to the third group – are both classified in the middle group according to the SUNflowerNext research. These differences can be partly explained by looking at the ordinal data presented in Table 10. The relatively high index score of Slovenia can be justified by the fact that this country never ranks worst than 13th on an indicator. The opposite is true for the low index score of Cyprus which scores extremely bad (19th, 20th, 20th and 21st) on four of the six indicators. The average index score of Ireland and Austria can be clarified by only average performances (i.e., the ranks of Austria are between 8 and 15.5) or a combination of very good and very bad performances (i.e., Ireland scores 1st and 2nd but also 18th twice). Finally, the average composite road safety performance index scores of Sweden and Finland are rather atypical (given their low number of road fatalities per million inhabitants; see Figure 14). Their worse roads score – or more generally, the selected roads indicator – plays a part in this.

Next, the road safety performance index is compared to the final outcome indicator, the human development index and the corruption perceptions index. The Spearman's correlation coefficient and the Wilcoxon matched-pairs signed ranks test (see e.g., Sheskin, 1997) are used to evaluate the degree of consistency between the RSPI ranking and the FO, HDI and CPI ranking. All three correlation coefficients appeared to be significant at the 1% level. Moreover, an increase in human development index score (representing a long and healthy life, knowledge and a decent standard of living) is associated with an increase in road safety performance index score. At the same time, an increase in CPI score – i.e., moving to less corruption or more cleanliness – corresponds to an increase in road safety performance. Based on the results of the Wilcoxon test (i.e., test scores smaller than the critical value) it could be concluded that the countries do not perform significantly different with respect to RSPI and FO respectively HDI and CPI. In general, it can be concluded that the developed road safety performance index has a link with the final outcome indicator, the human development index and the corruption perceptions index. In the future, more indexes and relationships could be studied in order to describe country profiles related to various aspects of society.

This research is concluded with some final visualisations. Apart from the coloured map (Figure 27) showing the countries with a high, average and low road safety performance, a graph with detailed index information is created per country. More specifically, two pie charts are shown for each country. In Figure 28, this is illustrated for the first country in the data set. Appendix III shows the charts for all 21 countries in decreasing order of index score starting with the Netherlands (0.692) and ending with Cyprus and Estonia (0.000). The light grey area in the left pie chart represents the size of the index score. In addition to the overall index score the division into its components is presented by the pie chart at the right-hand side. Each indicator share is the product of the rescaled indicator value and the corresponding weight. Starting at the top and moving clockwise, the relative share of alcohol and drugs (dark blue), speed (red), protective systems (green), vehicle (purple), roads (light blue) and trauma management (orange) in the overall index score is shown. By presenting these shares it is specified to which extent the index score of a particular country relates to the six domains. The size of the shares depends on the indicator value and a specific weight both relating to a certain ordered position and respecting the limits imposed by experts based on budget allocation.

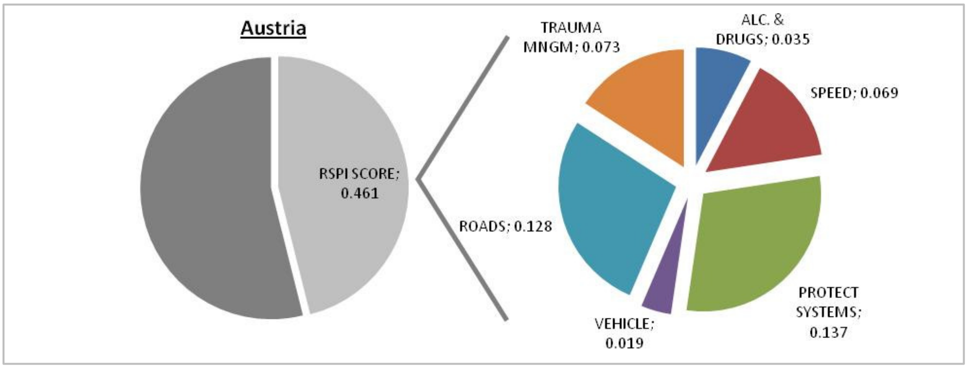


Figure 28: Visualisation of the road safety performance index score and its components for Austria

Finally, the 21 countries are graphed based on their road safety performance index score (x-axis) and their final outcome value (y-axis).

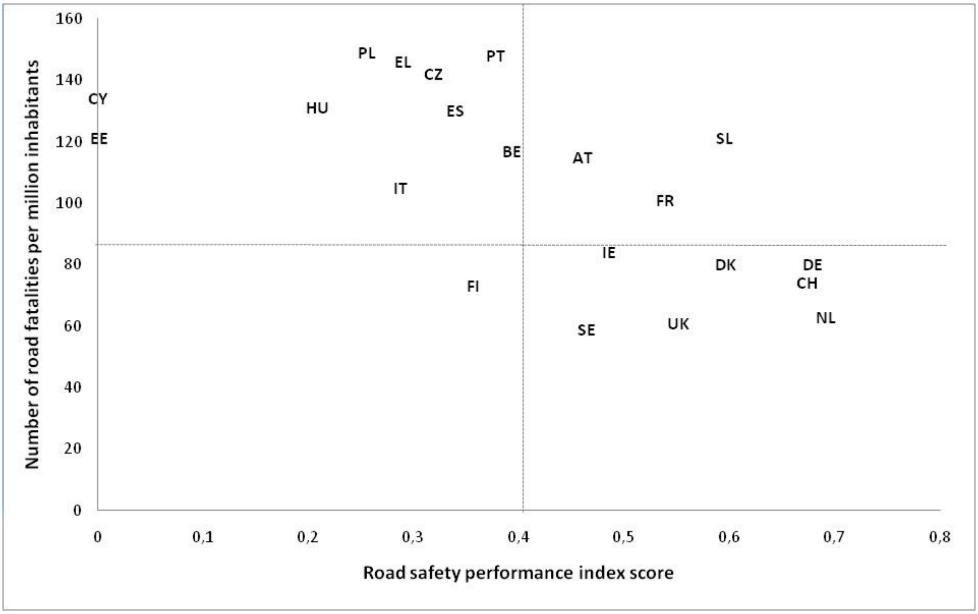


Figure 29: Road safety performance index scores versus number of road fatalities per million inhabitants

In general, it can be seen from Figure 29 that – as expected – a higher RSPI score corresponds to a lower FO value and vice versa. Moreover, by dividing the graph space into four quadrants it becomes clear that for most countries this negative association is true (i.e., the countries situated in the top left and bottom right quadrant). However, for four countries the road safety performance index score gives a bad indication of the number of road fatalities per million inhabitants. The data set used might be a cause hereof. The risk domains alcohol and drugs, speed, protective systems, vehicle, roads and trauma management were represented by one best available indicator. Austria, France and Slovenia have no bottom-five performances whereas Finland has two (on the roads and trauma management indicator).

10.4 Concluding remarks

This chapter discussed the last step in the construction of the road safety performance index: the computation, evaluation and visualisation of the final index scores. First, the final index methodology was decided upon taking all previously acquired information into account. More specifically, the most promising weighting method – data envelopment analysis – was further elaborated in terms of aggregation. The most optimal index score under the imposed restrictions was computed for each country. This index score was at the same time an acceptable solution due to the incorporation of weight boundaries regarding the share of each indicator in the index score (deduced from budget allocating experts) and the emphasizing of bad performances (by means of an orness restriction). Applying the suggested index methodology on the best available data set (and keeping the limitations in mind) three groups of countries could be distinguished. The Netherlands, Germany, Switzerland, Denmark, Slovenia, United Kingdom and France obtained a high road safety performance index score; Ireland, Sweden, Austria, Belgium, Portugal, Finland and Spain had an average index score; and Czech Republic, Greece, Italy, Poland, Hungary, Cyprus and Estonia were characterized by a low index score.

Next, the relationship between the created road safety performance index and related indexes and indicators was investigated. Using the Spearman's correlation coefficient and the Wilcoxon test a high degree of consistency between the countries' ranking based on the road safety performance index on the one hand and the final outcome, human development index and corruption perceptions index ranking on the other hand was found. The direction of the relationships was as expected. A higher human development index score can be associated with a better road safety performance. A more clean (i.e., less corrupt) perceived country is on average characterized by a higher road safety performance index score. Finally, the number of road fatalities per million inhabitants is negatively associated with the road safety performance index. It can be concluded that research on the intermediate outcome level is promising and relevant as it might provide signals that action is needed. The combination of road safety performance indicators in an index using a scientifically sound and appropriate methodology provides valuable insights and offers new perspectives within future road safety policy making. The final chapter discusses the main conclusions of this manuscript and some directions for further research.

Chapter 11 Final conclusions

In this final chapter, key findings (Section 11.1) as well as directions for future research (Section 11.2) are discussed.

11.1 Key findings

Four key findings can be formulated regarding the research described in this manuscript:

1. Comparing the performance of countries in terms of road safety by means of an index at the intermediate outcome level enables earlier and goal-oriented action
2. Among the methodological options involved in the index construction process the ones offering additional insights and resulting in a clear link with the final outcome ranking should be preferred
3. Final – optimal – index scores are obtained from a data envelopment analysis model with orness restriction guaranteeing country-specific yet acceptable weights applying to the meaning of an indicator as well as to the magnitude of the performance
4. Currently, the degree of maturity of road safety performance indicators and the availability and quality of indicator data highly varies between the risk domains giving rise to data opportunities in the future

Each finding is described successively.

1. Comparing the performance of countries in terms of road safety by means of an index at the intermediate outcome level enables earlier and goal-oriented action

Road safety is an important policy area that can benefit from the use of indicators and indexes. In this research, the complex and multidisciplinary concept of road safety is translated into quantifiable indicators. Apart from the currently used final outcome indicators – such as the number of road fatalities per million inhabitants – safety performance indicators situated on the level of intermediate outcomes are studied. These indicators represent important road safety risk domains and help in detecting the key problem areas in a particular country. That way, road safety enhancing measures can be taken before the risk would manifest itself in an increased number of accidents and casualties.

In this research, we start from the seven risk domains used in the European SafetyNet project (Vis, 2005). Safety performance indicators are identified with respect to alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management. Countries could be compared on each safety performance indicator individually. However, a composite road safety performance index resulting from the combination of a set of indicators presents an enriched picture of road safety performance. A multitude of risk information is captured in a concise and comprehensive way. Moreover, the index scores enable an easy comparison of road safety performance across countries.

In other research fields, the idea of an index has already been elaborated (see e.g., Saisana and Tarantola, 2002). As the concept of road safety performance indicators is relatively new, very limited research effort has been put in road safety index research. In order to become a valuable tool a scientifically sound and appropriate index methodology (see also key findings 2 and 3) and appropriate safety performance indicators (see also key finding 4) are required.

Contrary to the traditional road safety research, the relative road safety performance of countries is assessed by means of a composite road safety performance index representing several road safety performance aspects rather than only the number of road fatalities per million inhabitants. Nevertheless, the link between the intermediate outcome level – i.e., the composite road safety performance index – and the final outcome level – i.e., the number of road fatalities per million inhabitants – is of interest in this research. The level of agreement between the two demonstrated that research on the intermediate outcome level is promising and relevant as the road safety performance index and its indicators can act as warning signals for urgent action.

2. Among the methodological options involved in the index construction process the ones offering additional insights and resulting in a clear link with the final outcome ranking should be preferred

To obtain an index a number of steps which are in fact research domains in itself need to be gone through: selecting appropriate indicators to combine, collecting indicator data, performing data analysis, weighting and aggregating the individual indicators, testing the robustness of the road safety index and computing, evaluating and presenting the final index scores. The road safety performance index can be constructed in various ways. Nevertheless, a scientifically sound and appropriate methodology is a prerequisite for its acceptance and use. After all, the index information will be used for ranking countries based on their overall road safety performance, for monitoring the evolution in road safety risk and for supporting policy decisions. Therefore, each step in this process has been studied and different options have been evaluated with respect to the road safety index case. The main methodological challenges were related to the indicator weights, the aggregation process and the robustness testing.

The determination of indicator weights based on five common methods was discussed. Weights obtained by means of factor analysis are based on correlations; analytic hierarchy process and budget allocation weights result from experts' opinions questioned in a different way; data envelopment analysis weights are obtained from a constrained optimization model; equal weighting assigns the same weight to each indicator. The main advantages, disadvantages and requirements of each method were described. Subsequently, each method was applied and the countries' rankings based on the computed index scores were compared. A qualitative assessment of the five methods in addition to a quantitative one – based on the correlation coefficient between each index ranking and the final outcome ranking – disclosed that the data envelopment analysis method is the most promising weighting method of the five in the road safety index context. The DEA model strongly weighted the behavioural indicators alcohol and drugs, speed and protective systems and assigned very limited weight to the roads indicator.

The process of aggregating indicators into an index was also studied. Considering the different types of aggregation operators, the class of averaging operators – in which the index score is bounded by the lowest and highest indicator value – and in particular, the ordered weighted averaging (OWA) operators – assigning weights to magnitudes of performances – appeared to be the most useful aggregation operators for the road safety index case. Special types of OWA operators (e.g., the arithmetic mean operator) were discussed and different weighting vectors (corresponding to a specific α value) interpreted. Moreover, the attitude of decision makers or experts could easily be incorporated in the aggregation process thereby enhancing the value and acceptability of the constructed index. Linguistic aggregation opinions were transformed into restrictions that α should satisfy. The selected OWA operator (with α equal to two) implies that the index takes all indicators into account but emphasizes worse performances. In other words, the index score reflects a rather intolerant aggregation behaviour in which compensation of bad scores by good ones is only limitedly allowed.

Finally, the robustness of the road safety performance index deserves particular attention. Given the multistage index process, the index developer is left with a number of methodological decisions that might influence the final countries' ranking. Performing uncertainty and sensitivity analyses is an essential part of the index process as it shows how robust the index is and which extra information would imply more robustness. The impact of a change in indicator selection, normalisation technique, weighting method, expert selection and way of aggregating on the absolute average shift in countries' ranking with respect to a reference ranking was quantified. The first reference ranking – the default index ranking – was used to assess the impact of a change in methodology compared to the countries' ranking using the most simple or common methods (i.e., the linear aggregation of six equally weighted standardised indicator values). A second reference ranking – the final outcome ranking – was used in order to assess the degree of agreement with the ranking based on the number of road fatalities per million inhabitants over the possible methodological scenarios. The absolute average shift in rank appeared to be 1.96 positions with respect to the default index ranking and 3.87 positions with respect to the final outcome ranking. By means of the sensitivity indexes it was concluded that the weighting method and the indicator selection are crucial for obtaining robust results and therefore require careful evaluation and justification.

In general, it can be concluded that insight was gained in the index construction process. The different methodological steps (such as weighting) were thoroughly examined and the various options assessed in a quantitative and qualitative way.

3. Final – optimal – index scores are obtained from a data envelopment analysis model with orness restriction guaranteeing country-specific yet acceptable weights applying to the meaning of an indicator as well as to the magnitude of the performance

Besides exploring the different options involved in the index process, a final index methodology was specified taking the previously acquired information into account. The robustness testing demonstrated the impact of the weighting method and the indicator selection. The composite road safety performance index was chosen to combine six indicators identified as best available road safety performance indicators for the alcohol and drugs, speed, protective systems, vehicle, roads and trauma management domain. Moreover, a final decision upon the weighting method had to be made. In the current context, the data envelopment analysis method is preferable to the other four methods. The indicator weights can be justified as they are a combination of objective and subjective information. On the one hand, indicator weights are derived from the data; on the other hand, acceptable weights are obtained as prior knowledge or accepted views from road safety experts can be easily included as boundaries. The constrained optimization model results in the most optimal index score for each country.

After the selection of the indicators and the data envelopment analysis weighting method, the other methodological aspects needed justification. In particular, the linguistic formulations with respect to aggregation were taken into account. The intolerant aggregation behaviour was incorporated in the DEA model by adding an orness restriction. This extra restriction entailed the use of ordered values. The data were rescaled to deduce the sequence of performance for each country.

In the end, an optimization model in which the indicator weights respect the restrictions imposed by experts concerning the share of each indicator in the overall index score as well as the aggregation idea of emphasizing bad performances was proposed as a scientifically sound and appropriate methodology for the creation of a road safety performance index consisting of six best available indicator values. For each country, a composite index value was obtained. Three groups of countries could be distinguished. The first group, characterized by a good road safety performance, consists of the Netherlands, Germany, Switzerland, Denmark, Slovenia, United Kingdom and France. Secondly, Ireland, Sweden, Austria, Belgium, Portugal, Finland and Spain have an average road safety performance. Thirdly, Czech Republic, Greece, Italy, Poland, Hungary, Cyprus and Estonia are characterized by a bad road safety performance.

4. Currently, the degree of maturity of road safety performance indicators and the availability and quality of indicator data highly varies between the risk domains giving rise to data opportunities in the future

This manuscript illustrated the index methodology using a set of best available indicator values for 21 European countries relating to 2003 (i.e., the most recent year for which values were available). For the seven risk domains used in the European SafetyNet project – alcohol and drugs, speed, protective systems, daytime running lights, vehicle, roads and trauma management – possible road safety indicators were listed from literature and subsequently evaluated taking several selection criteria into account. That way, the best needed and best available indicator for each risk domain were identified. Among all possible indicators, the best needed indicator was the most relevant, measurable, interpretable, specific and sensitive one. However, the index methodology was illustrated on a set of best available indicators, i.e., indicators scoring best on these five criteria as well as with respect to data availability, reliability and comparability. Since the concept of road safety performance indicator is relatively new and comparable data were to be found for a large set of countries, data issues resulted in a discrepancy between the best needed and the best available indicator set.

The identification of appropriate indicators for the daytime running lights domain appeared to be problematic. As the best available daytime running lights indicator already scored below-average on the selection criteria it was decided to incorporate no daytime running lights indicator in the index at this moment. Moreover, the best available indicator for representing the roads domain and the trauma management domain appeared to be proxy indicators having a less clear link with the final outcome indicator than the other indicators did. In addition, self-reported data were used for the alcohol and drugs as well as the speed indicator. Therefore, the results should be interpreted in terms of the indicators and data used. However, the road safety performance index created here consists of six indicators which we believe reflect essential and best available information to track road safety performance. Based on the multivariate analyses it could be concluded that in general, the six best available indicators[47] are appropriate for combination in an index (i.e., the data set has a satisfactory level of correlation, internal consistency, clustering and explanatory power).

Furthermore, the methodology described in this manuscript can be applied to other data sets as well. The future gain in importance of concepts like road safety performance indicator and road safety performance index will give rise to new data opportunities. The index will benefit from the increase in data availability in the near future and better reflect the road safety performance of countries. At this point, the need for high-quality and comparable indicator data for a large set of countries over time cannot be overemphasized. The indicators should be produced with adequate frequency to support decision makers in a systematic way, i.e., for identifying problems, monitoring implementation of measures and following-up their effectiveness.

11.2 Topics for further research

The topic of a road safety performance index is an interesting one that will inspire future research. This manuscript aimed at making a valuable contribution to this by developing a scientifically sound and appropriate methodology for the creation of a composite road safety performance index. A first application of the index methodology on a recent road safety indicator data set (Wegman et al., 2008) already produced promising results (Shen et al., 2009). Below, three interesting challenges for future research are stated.

[47] i.e., the percentage of car drivers reporting to often drive while having a blood alcohol content below the legal limit (data from Social Attitudes to Road Traffic Risk in Europe), the percentage of car drivers reporting to often drive in built-up areas at a speed below the legal limit (Social Attitudes to Road Traffic Risk in Europe), the percentage of persons wearing a seat belt in the front seats of a car or van (European Transport Safety Council), the share of relatively new cars (i.e. less than 6 years old) (Eurostat), the density of motorways (Eurostat) and the share of the gross domestic product spent on health care (World Health Organization).

A. Performing index research based on more information

The index process has been explored and illustrated using six best available indicators. A first extension of this research is to further develop the existing road safety risk domains and the road safety performance indicators used for each risk domain. Additional risk domains may be worthwhile incorporating in the index. Moreover, a certain type of road safety risk would be more realistically captured in case several indicators were used. In that case, a layered index methodology needs to be applied. More specifically, the overall composite road safety performance index then joins several domain indexes which are combinations of indicators in itself. Although the acquired knowledge regarding indicator selection, data preparation, weighting and aggregation methods and robustness testing can be used, new challenges arise. The number of indicators needed in each domain and the most optimal hierarchy structure need to be determined.

In addition to extra safety performance indicators, other final outcome indicators could be incorporated in the analysis. The number of road fatalities per million inhabitants has been used in the current research because it makes us of the most comparable and reliable accident and exposure data. However, in the future, other final outcome indicators could become more appropriate (e.g., the number of road fatalities per million kilometres travelled) for relating the road safety performance index ranking to. Moreover, a distinction between safety performance indicators linked to the occurrence of an accident on the one hand and the severity of the injuries on the other might be worthwhile investigating. This would imply the creation of a separate accident frequency index and an accident severity index.

Thirdly, the scale on which the benchmarking occurs could be expanded. To offer a more complete picture, it would be interesting to increase the set of 21 European countries currently involved in the analysis. In case the relative degree of road safety performance is assessed on a larger (e.g., worldwide) scale a classification in comparable groups should be made first. The appropriateness of an indicator is among other things affected by the degree of development in mobility in the country (e.g., the seat belt wearing rate is less relevant in case a minority of trips occurs by motorised vehicles). Nevertheless, the index process remains the same and the best-in-class country could be assigned within each group based on the index scores.

Fourthly, road safety performance could be assessed over time. Due to data limitations, indicator values could only be obtained for 2003. Research on time series accident data is well-developed within the road safety field. In the same way, the monitoring of safety performance data collected at regular intervals would help in gaining a clear understanding of trends and expected progress towards postulated targets and benchmarks' performances.

To meet these research opportunities extra data in easy accessible sources are indispensable. Yet, an extension of the data set used for index research raises the issue of missing data. In order to take as much of the available information into account, we should aim to impute the missing values. Given its impact, a justifiable method needs to be applied. Therefore, future research attention will be paid to this.

B. Elaborating relevant road safety frameworks

A second direction for further research is amplifying on the road safety frameworks presented in the first chapter of this manuscript, i.e., the target hierarchy for road safety (Section 1.5.2) and the causal road safety framework (Section 1.5.1).

The intermediate outcome layer of the target hierarchy for road safety has been focused on in this research. Nevertheless, the different methodological steps involved in the creation of a composite road safety performance index could be applied to other types of indicators (i.e., other layers of the target hierarchy) as well. At the same time, it should be examined whether the combination of indicators on each layer of the pyramid requires a specific methodological approach. Moreover, is the creation of one overall index summarizing over the different layers of the pyramid relevant? Besides a vertical elaboration, several disaggregated road safety performance indexes could be developed, i.e., for certain age classes (e.g., elderly persons), transport modes (e.g., pedestrians), regions (e.g., NUTS) or transport means (e.g., sea transport).

The causal road safety framework consisted of a chain of eight components: driving forces, pressure, state, exposure, accident risk, injury risk, effects and actions. By identifying relevant indicators and collecting data the relationship between road safety performance and demographic, economic and geographic characteristics as well as cultural and social norms and the degree of exposure could be studied in detail.

The current research already found a high degree of consistency between the road safety performance index on the one hand and the human development index and the corruption perceptions index on the other. In the long term, it could be valuable to describe a country's profile by means of a number of diverse indexes (e.g., a road safety index, a technology index, a sustainability index, etc) in which a set of index scores describes the main characteristics of the country.

C. Coping with methodological challenges

Based on the research described in this manuscript a number of methodological ideas can be suggested for further research. First, the input-output data envelopment analysis model (described in Chapter 7) could be adapted in order to maximize the reduction in final outcomes while taking the required investment in measures into account. Moreover, the elaboration of a data envelopment analysis model over time would provide new insights because the degree of progress towards targets as well as the change regarding the benchmarks could be quantified.

Next, research with respect to the aggregation of indicators could focus on the creation of an index score in which interaction effects present between the indicators are incorporated. Finally, more attention could be paid to the sensitivity of some aspects. More specifically, one could benefit from gaining a clear insight into the impact of correcting extreme indicator values, omitting a few countries or a change in model restrictions.

Samenvatting

In dit onderzoek worden landen vergeleken op vlak van verkeersveiligheid. Het complexe en multidisciplinaire concept 'verkeersveiligheid' wordt hiervoor vertaald in kwantificeerbare indicatoren. Totnogtoe gebeurden vergelijkingen tussen landen voornamelijk op basis van geregistreerde ongevallendata (dus uiteindelijke uitkomsten). We willen echter inzicht krijgen in de belangrijkste onderliggende risicoaspecten van verkeersveiligheid. Op die manier kunnen gerichte maatregelen genomen worden om de belangrijkste risicoaspecten aan te pakken vooraleer deze tot uiting komen in een verhoogd aantal ongevallen en slachtoffers. Daarom worden verkeersveiligheidprestatie-indicatoren – gesitueerd op het niveau van tussentijdse uitkomsten – hier bestudeerd. In overeenstemming met zeven belangrijke risicodomeinen geïdentificeerd op Europees niveau beschouwen we indicatoren gerelateerd aan alcohol en drugs, snelheid, beschermende uitrusting, motorvoertuigverlichting overdag, voertuig, infrastructuur en nazorg.

Om landen te vergelijken op hun totale verkeersveiligheidprestatie focussen we op de combinatie van indicatoren in een samengestelde index. In geval van een index wordt een veelheid aan informatie vervat in één score, hetgeen communicatie en vergelijking vergemakkelijkt. In tegenstelling tot andere onderzoeksdomeinen zijn de geleverde onderzoeksinspanningen omtrent het opstellen van een verkeersveiligheidindex zeer beperkt. De hoofddoelstelling van dit manuscript is een wetenschappelijk correcte en geschikte methodologie te ontwikkelen voor het opstellen van een samengestelde verkeersveiligheidprestatie-index die kan gebruikt worden voor het monitoren en vergelijken van de verkeersveiligheidprestatie tussen landen. De verschillende methodologische aspecten verbonden aan het creëren van een index worden theoretisch toegelicht en geïllustreerd aan de hand van een set van indicatorwaarden. Op basis van de verkeersveiligheidprestatie-indexscore – waarin de prestatie op verschillende risicodomeinen vervat zit in plaats van enkel het aantal verkeersdoden – kunnen nuttige inzichten verkregen worden in de relatieve verkeersveiligheidprestatie van een set van Europese landen.

Na het introduceren van het concept 'indicator' en het schetsen van het verkeersveiligheidsprobleem geeft dit manuscript een overzicht van de verschillende methodologische stappen in het indexproces dewelke achtereenvolgens in detail aan bod komen. De belangrijkste aspecten worden hieronder kort toegelicht.

1. <u>Selecteren van geschikte indicatoren voor combinatie in een index</u>: voor elk risicodomein werden mogelijke indicatoren opgelijst op basis van literatuur en vervolgens geëvalueerd aan de hand van verschillende selectiecriteria. Elke indicator werd beoordeeld op vlak van relevantie, meetbaarheid, interpreteerbaarheid, specificiteit en gevoeligheid. Dit leidde tot de identificatie van ideale of aanbevolen indicatoren. Om de indexmethodologie te illustreren, werden echter drie extra criteria in rekening gebracht, namelijk data beschikbaarheid, betrouwbaarheid en vergelijkbaarheid. De opgestelde index bevatte zogenaamde best beschikbare indicatoren.
2. <u>Verzamelen van indicatordata</u>: verschillende databronnen werden geconsulteerd om gegevens te verzamelen voor de indicatoren geselecteerd in de vorige stap. Er werd getracht een dataset op te stellen met verkeersveiligheidprestatie-indicatorwaarden voor een groot aantal landen, met betrekking tot een bepaalde periode of (bij voorkeur zo recent mogelijk) jaar. De mate waarin prestatie-indicatoren ontwikkeld waren alsook het niveau van databeschikbaarheid en -kwaliteit varieerde tussen de risicodomeinen. Uiteindelijk werd voor zes best beschikbare indicatoren 2003 data voor 21 Europese landen bekomen ter kwantificering van de prestatie op vlak van verkeersveiligheid. De richting van de indicatoren werd op elkaar afgestemd waarbij hoge waarden (oftewel een betere verkeersveiligheidprestatie) na te streven zijn.
3. <u>Uitvoeren van data-analyses</u>: op de verzamelde dataset werden een aantal analyses uitgevoerd om meer inzicht te krijgen in elke indicator afzonderlijk (univariate analyse) alsook in de structuur en onderlinge verbanden in de gehele dataset (multivariate analyse). Voor elke indicator werden een aantal beschrijvende statistieken gegeven; normalisatietechnieken met als doel de data vergelijkbaar te maken over de indicatoren

werden besproken; en theoretische beschouwingen met betrekking tot ontbrekende waarden kwamen aan bod. Op basis van de resultaten van de multivariate analyse werd de geschiktheid van de indicatorset voor combinatie in een index aangetoond (i.e., voldoende correlatie, interne consistentie, clustering en verklarende kracht).

4. Een gewicht toekennen aan elke indicator: de werkwijze van vijf gangbare wegingsmethoden werd beschreven samen met hun voordelen, nadelen en vereisten. 'Factor analysis' gewichten zijn gebaseerd op correlaties; 'analytic hierarchy process' en 'budget allocation' gewichten resulteren uit expertmeningen bevraagd op een specifieke manier (i.e., de relatieve bijdrage per paar van indicatoren respectievelijk de verdeling van een budget); 'data envelopment analysis' (DEA) gewichten worden afgeleid uit een optimalisatiemodel met beperkingen; en 'equal weighting' kent een gewicht van 1/6 toe aan iedere indicator. Elke methode werd toegepast, de indexscores berekend en de uiteindelijke rangschikking van landen vergeleken. Op basis van een kwalitatieve beoordeling van de vijf methodes (in termen van interessante inzichten) evenals een kwantitatieve (op basis van de correlatiecoëfficiënt tussen elke indexrangschikking en de rangschikking op basis van het aantal doden per miljoen inwoners) werd de DEA methode verkozen.

5. Aggregeren van indicatoren: de wiskundige formule voor het combineren van indicatoren in een index moet bepaald worden. Naast de vaak toegepaste lineaire en geometrische aggregatie werd het onderzoeksdomein aggregatie als uitgangspunt genomen. De klasse van 'averaging operators' in het algemeen – waarbij de indexscore begrensd wordt door de laagste en hoogste indicatorwaarde – en de 'ordered weighted averaging' (OWA) operatoren in het bijzonder – waarbij gewichten worden toegekend aan groottes van prestaties – bleken waardevolle aggregatieoperatoren te zijn voor de verkeersveiligheidindex. Speciale gevallen van OWA operatoren (zoals de maximum operator) werden besproken en verschillende OWA wegingsvectoren geëvalueerd. De uiteindelijke vector weerspiegelde een eerder intolerant

aggregatie-idee waarin alle prestaties worden beschouwd, maar slechte prestaties benadrukt worden.

6. Testen van de robuustheid van de index: de verschillende stappen en opties in het indexproces impliceren dat de indexontwikkelaar een aantal methodologische keuzes moet maken die de uiteindelijke rangschikking van landen kunnen beïnvloeden. Het uitvoeren van een onzekerheid- en sensitiviteitanalyse is een essentieel onderdeel van het indexproces aangezien deze aantoont hoe robuust de index is en aangeeft welke extra informatie tot meer robuustheid kan leiden. De invloed van methodologische keuzes gemaakt bij het opstellen van de index – aangaande indicatorselectie, normalisatietechniek, wegingsmethode, expertkeuze en aggregatiewijze – op de absolute gemiddelde verschuiving in rangschikking op basis van de indexscores ten opzichte van een referentierangschikking werd gekwantificeerd. Daarnaast kon op basis van sensitiviteitindices besloten worden dat de wegingsmethode en de indicatorselectie een grote impact hebben en bijgevolg zorgvuldige evaluatie en verantwoording vereisen.
7. Berekenen, evalueren en visualiseren van de uiteindelijke indexscores: op basis van alle voorheen vergaarde informatie werd een uiteindelijke indexmethodologie voorgesteld. Meer bepaald werd de beste wegingsmethode – ‘data envelopment analysis’ – verder uitgewerkt met betrekking tot aggregatie. Voor elk land bepaalden we de meest optimale indexscore onder de vooropgestelde beperkingen. Deze indexscore was tevens een aanvaardbare oplossing omwille van de opgenomen beperkingen omtrent het aandeel van elke indicator in de indexscore (bepaald op basis van ‘budget allocation’) en het benadrukken van slechte prestaties (door middel van een ‘orness’ beperking). Het toepassen van de voorgestelde indexmethodologie op de best beschikbare dataset (en de beperkingen in het achterhoofd houdend) resulteerde in drie groepen van landen. Nederland, Duitsland, Zwitserland, Denemarken, Slovenië, het Verenigd Koninkrijk en Frankrijk behaalden een hoge verkeersveiligheidprestatie-indexscore;

Ierland, Zweden, Oostenrijk, België, Portugal, Finland en Spanje hadden een gemiddelde indexscore; en Tsjechië, Griekenland, Italië, Polen, Hongarije, Cyprus en Estland werden gekenmerkt door een lage indexscore. Verder werd de relatie tussen de opgestelde index en andere, gerelateerde indicatoren en indexen bepaald. Er bleek een hoge mate van samenhang te zijn tussen de rangschikking op basis van de verkeersveiligheidprestatie-index enerzijds en de 'final outcome', 'human development index' en 'corruption perceptions index' rangschikking anderzijds. Het indexproces werd afgesloten met een visualisatie van de uiteindelijke resultaten.

Over het algemeen kunnen we besluiten dat inzicht werd verkregen in het proces om een index te creëren. Elke stap in dit proces werd nauwgezet onderzocht. Uit de literatuur werden relevante opties voor de verkeersveiligheidindex geselecteerd en vervolgens grondig geëvalueerd. Kwalitatieve en kwantitatieve beoordeling resulteerde uiteindelijk in een wetenschappelijk correcte en geschikte indexmethodologie voor de combinatie van best beschikbare verkeersveiligheidprestatie-indicatoren in een samengestelde index. De meest optimale indexscore voor elk land werd berekend waarbij aan de opgelegde beperkingen aangaande het relatieve belang van een indicator en de relatieve grootte van een prestatie werd tegemoetgekomen.

Gedurende dit onderzoek werd de link tussen de verkeersveiligheidprestatie-index (het niveau van tussentijdse uitkomsten) en het aantal verkeersdoden per miljoen inwoners (het niveau van uiteindelijke uitkomsten) in gedachten gehouden. De mate van overeenstemming tussen de twee toont aan dat onderzoek op het niveau van tussentijdse uitkomsten veelbelovend en relevant is aangezien de verkeersveiligheidprestatie-index en zijn indicatoren dienst kunnen doen als waarschuwingssignaal en eerdere, specifieke actie toelaten. Naast het bespreken van het indexproces werd er een hoofdstuk gewijd aan het bepalen van een verkeersveiligheidscore die zowel indicatoren op het niveau van tussentijdse uitkomsten als op het niveau van uiteindelijke uitkomsten beschouwt. De ontwikkeling van een input-output DEA model liet een rangschikking van landen toe op basis van een score die verschillende types van verkeersveiligheidsindicatoren vervat. Door de kenmerken van een land in rekening te brengen, konden specifieke aanbevelingen ter verbetering van de verkeersveiligheid in dat land gedaan worden zoals het identificeren van voorbeeldlanden en het toekennen van prioriteiten.

Tot slot, een samengestelde verkeersveiligheidprestatie-index is een waardevol communicatie-, monitoring- en beleidsondersteunend instrument voor het vergelijken van de prestatie van landen op vlak van verkeersveiligheid. Een wetenschappelijk correcte en geschikte methodologie evenals goede verkeersveiligheidprestatie-indicatoren zijn een vereiste hierbij. De toekomstige toename in belang van concepten zoals verkeersveiligheidprestatie-indicator en verkeersveiligheidprestatie-index zullen aanleiding geven tot nieuwe data-opportuniteiten terwijl het onderzoek beschreven in dit manuscript een waardevolle bijdrage kan betekenen voor de ontwikkeling van verkeersveiligheidindexen in de toekomst.

References

Al Haji, G. (2003). *Road safety development index*. Presented at the 16th ICTCT workshop, Soesterberg.

Al Haji, G. (2005). *Towards a road safety development index*. PhD Thesis. Linköpings universitet.

Allen, R., Athanassopoulos, A., Dyson, R.G. and Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. Annals of Operations Research, 73 pp. 13-34.

Allen, R. and Thanassoulis, E. (2004). Improving envelopment in data envelopment analysis. European Journal of Operational Research, 154 (2) pp. 363-379.

Anderson, T. (2006). *A data envelopment analysis home page*. Retrieved August, 25th, 2006 from http://www.etm.pdx.edu/dea/homedea.html.

Balkin, S. and Ord, J.K. (2001). Assessing the impact of speed-limit increases on fatal interstate crashes. Journal of Transportation and Statistics, 4 (1) pp. 1-26.

Barr, R.S. (2004). *DEA software tools and technology: A state-of-the-art survey*. Southern Methodist University.

Begeleidingscomité (2001). *Staten-generaal van de verkeersveiligheid*.

Belgian Toxicology and Trauma Study Research Group (1997). *Een wetenschappelijk onderzoek over de aanwezigheid van alcohol, geneesmiddelen en illegale drugs bij slachtoffers van verkeersongevallen en het verband tussen deze stoffen en de ongevallen*.

Beliakov, G., Pradera, A. and Calvo, T. (2007). Aggregation functions: A guide for practitioners. Berlin: Springer.

Bird, S.M., Cox, D., Farewell, V.T., Goldstein, H., Holt, T. and Smith, P.C. (2005). Performance indicators: Good, bad, and ugly. Journal of the Royal Statistical Society: Series A, 168 (1) pp. 1-27.

Booysen, F. (2002). An overview and evaluation of composite indices of development. Social Indicators Research, 59 pp. 115-151.

Cambridge Systematics Inc. and Meyer, M.D. (2008). *Crashes vs. congestion: What's the cost to society?*

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2 pp. 429-444.

Cherchye, L. and Kuosmanen, T. (2004). *Benchmarking sustainable development: A synthetic meta-index approach*. United Nations University - World Institute for Development Economics Research.

Cherchye, L., Moesen, W. and Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the EU. Journal of Common Market Studies, 42 (5) pp. 919-955.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R. and Tarantola, S. (2006). *Creating composite indicators with DEA and robustness analysis: The case of the technology achievement index*. Catholic University of Leuven and Joint Research Centre.

Chiang, C.-M. and Lai, C.-M. (2002). A Study on the comprehensive indicator of indoor environment assessment for occupants' health in Taiwan. Building and Environment, 37 (4) pp. 387-392.

Christens, P.F. (2003). *Statistical modelling of traffic safety development*. PhD Thesis. Technical university of Denmark.

Cooper, W.W., Seiford, L.M. and Tone, K. (2000). Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software. Dordrecht: Kluwer Academic Publishers.

Cooper, W.W., Seiford, L.M., Thanassoulis, E. and Zanakis, S.H. (2004). DEA and its uses in different countries. European Journal of Operational Research, 154 (2) pp. 337-344.

De Brabander, B. and Vereeck, L. (2007). Valuing the prevention of road accidents in Belgium. Transport Reviews, 27 (6) pp. 715-732.

Despontin, M., de Brucker, K., Coeck, C. and Verbeke, A. (1998). *The economic evaluation of road safety in the European Union*. European Union.

Directorate for Science Technology and Industry (2003). *Composite indicators of country performance: A critical assessment*. Organization for Economic Co-operation and Development.

Dupont, F. (2005). *Instruments with strong policy impact require strong methodology: Example of the European Innovation Scoreboard 2005*. Presented at the 2006 Workshop on European Composite Indicators and Scoreboards, Brussels.

Economic Commission for Europe (2008). *Annual bulletin of transport statistics for Europe and North America*. United Nations.

Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident Analysis and Prevention, 36 pp. 637-647.

Elvik, R. (1993). The effects on accidents of compulsory use of daytime running lights for cars in Norway. Accident Analysis and Prevention, 25 pp. 383-398.

Elvik, R. and Vaa, T. (2004). Handbook of road safety measures. Oxford: Elsevier Ltd.

Elvik, R. (2005). *Speed and road safety: Synthesis of evidence from evaluation studies*. Presented at the 84th annual meeting of the Transportation Research Board, Washington D.C.

Environment Canada, CMHC and Westland Resource Group (1999). *Sustainable community indicators software: Applications and proposed indicators*.

European Commission (2001). *White paper European transport policy for 2010: Time to decide*. Commission of the European Communities.

European Commission (2004a). *COST Action 329: Models for traffic and safety development and interventions*. Commission of the European Communities.

European Commission (2004b). *European innovation scoreboard 2004: Comparative analysis of innovation performance*. Commission of the European Communities.

European Commission (2005). *Sustainable development indicators to monitor the implementation of the EU sustainable development strategy*. Commission of the European Communities.

European Commission (2006a). *European road safety action programme - Mid-term review*. Commission of the European Communities.

European Commission (2006b). *Saving lives with daytime running lights (DRL): A consultation paper presented by the Directorate General for Energy and Transport*. Commission of the European Communities.

European Environment Agency (1999). *Environmental indicators: Typology and overview*. European Environment Agency.

European Environment Agency and Eurostat (1999). *Towards a transport and environment reporting mechanism (TERM) for the EU. Part 2: Some preliminary indicator sheets*.

European Road Safety Observatory (2006). *Quantitative road safety targets*.

European Transport Safety Council (1999). *Reducing the severity of road injuries through post impact care*. European Transport Safety Council.

European Transport Safety Council (2001). *Transport safety performance indicators*. European Transport Safety Council.

European Union Road Federation (2006). *European Road Statistics 2006*.

European Union Road Federation (2007). *European Road Statistics 2007*.

European Union Road Federation (2008). *European Road Statistics 2008*.

Eurostat (2007). *Europe in figures: Eurostat yearbook 2006-07*. Commission of the European Communities.

Eurostat (2008). *Transport database*. Retrieved March, 4th, 2008 from http://epp.eurostat.ec.europa.eu.

Evans, L. (2004). Traffic safety. Michigan: Bloomfield Hills.

Expert Choice Inc. (2006). Expert Choice 11.

Farchi, S., Molino, N., Rossi, P.G., Borgia, P., Krzyzanowski, M., Dalbokova, D., Kim, R. and European Road Accident Indicator Working Group (2006). Defining a common set of indicators to monitor road accidents in the European Union. BMC Public Health, 6 pp. 183.

Federal Highway Administration (2004). *Transportation performance measures in Australia, Canada, Japan and New Zealand*. Federal Highway Administration.

Federale Commissie voor de Verkeersveiligheid (2007). *Staten-Generaal van de Verkeersveiligheid*.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. and Thomsen, L.K. (1995). Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. Accident Analysis and Prevention, 27 (1) pp. 1-20.

Grabisch, M., Orlovski, S.A. and Yager, R.R. (1999). Fuzzy aggregation of numerical preferences. In Fuzzy sets in decision analysis, operations research and statistics. USA: Kluwer Academic Publishers.

Gyomai, G. and Guidetti, E. (2008). *OECD Sytem of composite leading indicators*. Organisation for Economic Co-operation and Development.

Haas, R. and Meixner, O. (2006). *An illustrated guide to the analytic hierarchy process*. Institute of Marketing & Innovation - University of Natural Resources and Applied Life Sciences.

Haddon, W. (1980). Options for the prevention of motor vehicle crash injury. Israeli Medical Journal, 16 pp. 45-65.

Hakim, S., Shefer, D., Hakkert, A.S. and Hocherman, I. (1991). A critical review of macro models for road accidents. Accident Analysis and Prevention, 23 (5) pp. 379-400.

Hakkert, A.S. and Wesemann, P. (2005). *ROSEBUD WP3: The use of efficiency assessment tools: Solutions to barriers*.

Hakkert, A.S. and Gitelman, V., (Eds) (2007). *Road safety performance indicators: Manual*. D3.8 of the EU FP6 project SafetyNet.

Hakkert, A.S., Gitelman, V. and Vis, M.A., (Eds) (2007). *Road safety performance indicators: Theory*. D3.6 of the EU FP6 project SafetyNet.

Hens, L., Lafère, J. and De Wit, J. (2005). *Inventarisatie van milieu-gezondheids-indicatoren op internationaal en nationaal vlak*. Steunpunt Milieu en Gezondheid.

Hermans, E., Brijs, T., Stiers, T. and Offermans, C. (2006a). *The impact of weather conditions on road safety investigated on an hourly basis*. Presented at the 85th annual meeting of the Transportation Research Board, Washington D.C.

Hermans, E., Van den Bossche, F. and Wets, G. (2006b). Describing the evolution in the number of highway deaths by decomposition in exposure, accident risk and fatality risk. Transportation Research Record, 1950 pp. 1-8.

Hermans, E., Van den Bossche, F. and Wets, G. (2006c). The frequency and severity of road traffic accidents studied by state space methods. Journal of Transportation and Statistics, 9 (1) pp. 63-76.

Hermans, E., Ruan, D., Brijs, T., Wets, G. and Vanhoof, K. (2008a). *Evaluation of road safety performance indicators using OWA operators.* Presented at the 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid.

Hermans, E., Van den Bossche, F. and Wets, G. (2008b). Combining road safety information in a performance index. Accident Analysis and Prevention, (40) pp. 1337-1344.

Hermans, E., Brijs, T., Wets, G. and Vanhoof, K. (2009a). Benchmarking road safety: Lessons to learn from a data envelopment analysis. Accident Analysis and Prevention, 41 (1) pp. 174-182.

Hermans, E., Van den Bossche, F. and Wets, G. (2009b). Uncertainty assessment of the road safety index. Reliability Engineering and System Safety, 94 pp. 1220-1228.

Hussain, I.M. and Redmond, A.D. (1994). Are pre-hospital deaths from accidental injury preventable? British Medical Journal, 308 pp. 1077-1080.

Janssens, D. and Wets, G. (2005). *The presentation of an activity-based approach for surveying and modelling travel behaviour*. Presented at the 32nd Colloquium Vervoersplanologisch Speurwerk, Antwerp.

Johnson, R.A. and Wichern, D.W. (2002). Applied multivariate statistical analysis. New Jersey: Pearson Education.

Kaminsky, G., Lizondo, S. and Reinhart, C.M. (1998). *Leading indicators of currency crises*. International Monetary Fund.

Kaplan, R.S. and Norton, D.P. (1992). The balanced scorecard--measures that drive performance. Harvard Business Review, 70 (1) pp. 71-79.

Kleinknecht, A., Van Montfort, K. and Brouwer, E. (2002). The non-trivial choice between innovation indicators. Economics of Innovation and New Technology, 11 (2) pp. 109-121.

Knight, I., Sexton, B., Bartlett, R., Barlow, T., Latham, S. and McCrae, I. (2006). *Daytime running lights (DRL): A review of the reports from the European Commission*. Transport and Road Research Laboratory.

Koornstra, M.J., Lynam, D., Nilsson, G., Noordzij, P., Pettersson, H.E., Wegman, F. and Wouters, P. (2002). *SUNflower: A comparative study of the development of road safety in Sweden, the United Kingdom and the Netherlands*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.

Kweon, Y. and Kockelman, K.M. (2005). *The safety effects of speed limit changes: Use of panel models, including speed, use and design variables*. Presented at the 84th annual meeting of the Transportation Research Board, Washington D.C.

Lambsdorff, J.G. (2003). *Background paper to the 2003 corruption perceptions index*. Transparency International and University of Passau.

Lambsdorff, J.G. (2004). *Background paper to the 2004 corruption perceptions index*. Transparency International and University of Passau.

Lammar, P. (2006). *Verkeersveiligheidsindicatoren: Overzicht en omschrijving van bestaande en bruikbare indicatoren*. Steunpunt Verkeersveiligheid.

Land Transport Safety Authority (2000). *Working paper 6: Predicting and costing road safety outcomes*.

Ledoux, L., Mertens, R. and Wolff, P. (2005). EU sustainable development indicators: An overview. National Resources Forum, 29 (4) pp. 392-403.

LINDO Systems Inc. (2007). Optimization Modeling with LINGO. Version 10.0.

Litman, T. (2005). *Well measured: Developing indicators for comprehensive and sustainable transport planning*. Victoria Transport Policy Institute.

Litman, T. (2007). *Developing indicators for comprehensive and sustainable transport planning*. Presented at the 86th annual meeting of the Transportation Research Board, Washington D.C.

Little, R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data. New York: John Wiley & Sons.

Luukkanen, L. (2003). *Safety management system and transport safety performance indicators in Finland*. Likkenneturva.

Lynam, D., Hummel, T., Barker, J., Lawson, S.D. and Joanne, H. (2004). *EuroRAP Year 1 (2003) update - performance tracking of roads*.

Maclaren, V.W., Labatt, S., McKay, J. and Vande Vegte, M. (1995). *Developing indicators of urban sustainability: A focus on the Canadian experience*. Presented at the Measuring Urban Sustainability: Canadian Indicators Workshop, Toronto.

Malhotra, N.K. and Birks, D.F. (1999). Marketing research: An applied approach. London: Prentice Hall, Inc.

McArthur, J.W. and Sachs, J.D. (2001). *The growth competitiveness index: Measuring technological advancement and the stages of development*. Center for International Development at Harvard University.

Melinder, K. (2007). Socio-cultural characteristics of high versus low risk societies regarding road traffic safety. Safety Science, 45 pp. 397-414.

Ministerie van de Vlaamse Gemeenschap (2001). *Ontwerp mobiliteitsplan Vlaanderen*. Retrieved November, 11th, 2004 from http://www.mobiel.vlaanderen.be/mobiliteitsplan/mobiliteitsplan01.php.

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

Morsink, P., Oppe, S., Reurings, M. and Wegman, F. (2007). Development of a footprint methodology for road safety. Transportation Research Record, 2009 pp. 104-112.

Munda, G. and Nardo, M. (2005). *Non-compensatory composite indicators for ranking countries: A defensible setting*. Institute for the Protection and Security of the Citizen.

Nardo, M., Tarantola, S., Saltelli, A., Andropoulos, C., Buescher, R., Karageorgos, G., Latvala, A. and Noel, F. (2004). *The e-business readiness composite indicator for 2003: A pilot study*. Joint Research Centre.

Nardo, M., Saisana, M., Saltelli, A. and Tarantola, S. (2005a). *Key economy indicators. D7.1: State-of-the-art report on simulation and indicators*. Joint Research Centre.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A. and Giovannini, E. (2005b). *Handbook on constructing composite indicators: Methodology and user guide*. Organisation for Economic Co-operation and Development.

National Cooperative Highway Research Program (2003). *Performance measures of operational effectiveness for highway segments and systems*. Transportation Research Board.

National Research Council (2002). *Key transportation indicators: Summary of a workshop*. National Academy Press.

National Road Safety Committee (2000). *Road safety strategy 2010: A consultation document.* National Road Safety Committee.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). Applied linear statistical models. Boston: McGraw-Hill.

Nilsson, G. (2004). *Traffic safety dimensions and the power model to describe the effect of speed on accidents*. Lund Institute of Technology.

Noland, R.B. (2003). Medical treatment and traffic fatality reductions in industrialised countries. Accident Analysis and Prevention, 35 (6) pp. 877-883.

Odeck, J. (2005). Evaluating target achievements in the public sector: An application of a rare non-parametric DEA and Malmquist indices. Journal of Applied Economics, 8 (1) pp. 171-190.

Organisation for Economic Co-operation and Development (1997). *Road safety principles and models: Review of descriptive, predictive, risk and accident consequence models*. Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development (2001). *Performance indicators for the road sector: Summary of the field tests*. Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development (2002). *Safety on roads: What's the vision?* Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development (2003). *Composite indicators of country performance: A critical assessment*. Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development and European Conference of Ministers of Transport (2006). *Country reports on road safety performance*.

Organisation for Economic Co-operation and Development (2008). *OECD Factbook 2008*. Organisation for Economic Co-operation and Development.

Parada, M.A., Cohn, L.D., Gonzalez, E., Byrd, T. and Cortes, M. (2001). The validity of self-reported seatbelt use: Hispanic and non-Hispanic drivers in El Paso. Accident Analysis and Prevention, 33 pp. 139-143.

Pedraja-Chaparro, F., Salinas-Jimenez, J. and Smith, P. (1997). On the role of weight restrictions in data envelopment analysis. Journal of Productivity Analysis, 8 pp. 215-230.

Pennoni, F., Tarantola, S. and Latvala, A. (2005). *The 2005 European e-business readiness index*. European Commission DG Joint Research Centre.

Puolamaa, M., Kaplas, M. and Reinikainen, T. (1996). *Index of environmental friendliness: A methodological study*.

Raeside, R. and White, D. (2004). Predicting casualty numbers in Great Britain. Transportation Research Record, 1897 pp. 142-147.

Roodman, D. (2005). *Commitment to Development Index*. Presented at the 2006 Workshop on European composite indicators and scoreboards, Brussels.

Rumar, K. (1999). *Transport safety visions, targets and strategies: Beyond 2000*. European Transport Safety Council.

Saaty, T.L. (1980). The analytic hierarchy process. USA: McGraw-Hill.

Sabey, B.E. and Taylor, H. (1980). *The known risks we run: The highway*. Transport and Road Research Laboratory.

SafetyNet (2005a). *State of the art report on risk and exposure data*. (D2.1).

SafetyNet (2005b). *Final proposal of EuroRIS*. (D6.3).

Saisana, M. and Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. Joint Research Centre.

Saisana, M., Saltelli, A. and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Royal Statistical Society, A, 168 (2) pp. 307-323.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). Global sensitivity analysis: The primer. Chichester: Wiley.

Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Centre for the Study of Living Standards.

SARTRE 3 consortium (2004). *European drivers and road risk*. Inrets.

Schafer, A. and Graham, J.W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7 (2) pp. 147-177.

Schafer, J.L. (1997). Analysis of incomplete multivariate data. New York: CRC Press.

Schafer, J.L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model*. Retrieved May, 4th, 2006 from www.stat.psu.edu/~jls/misoftwa.html.

Schafer, J.L. (2006). *The multiple imputation FAQ page*. Retrieved May, 4th, 2006 from http://www.stat.psu.edu/~jls/mifaq.html.

Scuffham, P.A. (2003). Economic factors and traffic crashes in New Zealand. Applied Economics, 35 (2) pp. 179-188.

Segnestam, L. (2002). *Indicators of environment and sustainable development: Theories and practical experience*. The International Bank for Reconstruction and Development.

Sharma, S. (1996). Applied multivariate techniques. New York: John Wiley & Sons.

Sharpe, A. (2004). *Literature review of frameworks for macro-indicators*. Centre for the Study of Living Standards.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T. and Vanhoof, K. (2009). *Road safety performance evaluation based on a multiple layer data envelopment analysis model*. (In submission).

Sheskin, D.J. (1997). Handbook of parametric and nonparametric statistical procedures. USA: CRC Press.

SPSS Inc. (2007). SPSS 16.0 for Windows. Version 16.0.1.

Steering Committee for the Review of Commonwealth/State Service Provision (1997). *Data envelopment analysis: A technique for measuring the efficiency of government service delivery*.

Studiedienst van de Vlaamse Regering (2006). *VRIND 2006: Vlaamse Regionale Indicatoren*.

SUPREME (2007). *Best practices in road safety: Handbook for measures at the country level*.

SWOV (2008). *SWOV Fact sheet: Daytime running lights*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.

Talbert, M.L., Balci, O. and Nance, R.E. (1994). *Application of the analytical hierarchy process to complex system design evaluation*. Virginia Polytechnic Institute and State University.

Tarantola, S., Saisana, M. and Saltelli, A. (2002). *Internal market index 2002: Technical details of the methodology*. Joint Research Centre.

Tarantola, S., Liska, R., Saltelli, A., Leapman, N. and Grant, C. (2004). *The internal market index 2004*. European Commission DG Joint Research Centre.

Tarantola, S., Nardo, M., Saisana, M. and Gatelli, D. (2006). A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator. Reliability Engineering and System Safety, 91 pp. 1135-1141.

Tessmer, J. (1999). Comparing international crash statistics. Journal of Transportation and Statistics, 2 (2) pp. 159-166.

Thoresen, T., Fry, T., Heiman, L. and Cameron, M.H. (1992). *Linking economic activity, road safety countermeasures and other factors with the Victorian road toll*. Monash University.

United Nations (2001). *Human development report 2001*. United Nations.

United Nations (2005). *Human development report 2005*. United Nations.

van Buuren, S. (2005). *Multiple imputation outline*. Retrieved April, 19th, 2006 from http://www.multiple-imputation.com.

Van den Bossche, F. and Wets, G. (2003). *Macro models in traffic safety and the DRAG family: Literature review*. Steunpunt Verkeersveiligheid.

Van den Bossche, F., Wets, G. and Brijs, T. (2005). *The role of exposure in the analysis of road accidents: A Belgian case-study*. Presented at the 84th annual meeting of the Transportation Research Board, Washington D.C.

Van Reeth, W. and Vanongeval, L. (2005). *Natuurrapport 2005: Indicatoren*.

Vanlaar, W. (2005). *The enhanced methodology of the Belgian drink driving roadside survey: A milestone for Belgium*. Belgisch Instituut voor de Verkeersveiligheid.

Vis, M.A. and Van Gent, A.L., (Eds) (2007a). *Road safety performance indicators: Country comparisons*. D3.7a of the EU FP6 project SafetyNet.

Vis, M.A. and Van Gent, A.L., (Eds) (2007b). *Road safety performance indicators: Country profiles*. D3.7b of the EU FP6 project SafetyNet.

Vis, M.A., (Ed) (2005). *State of the art report on road safety performance indicators*. D3.1 of the EU FP6 project SafetyNet.

Vlaams Ministerie van Mobiliteit en Openbare Werken (2007). *Verkeersveiligheidsplan Vlaanderen*.

Wackernagel, M. and Rees, W. (1997). Perceptual and structural barriers to investing in natural capital: Economics from an ecological footprint perspective. Ecological Economics, 20 pp. 3-24.

Washington, S.P., Lord, D. and Persaud, B. (2009). *The use of expert panels in highway safety: A critique*. Presented at the 88th annual meeting of the Transportation Research Board, Washington D.C.

Wayman, J.C. (2003). *Multiple imputation for missing data: What is it and how can I use it?* Presented at the American Educational Research Association, Chicago.

Wegman, F., Commandeur, J., Doveh, E., Eksler, V., Gitelman, V., Hakkert, S., Lynam, D. and Oppe, S. (2008). *SUNflowerNext: Towards a composite road safety performance index*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.

White, W.T. (1986). Does vehicle inspection prevent accidents? Accident Analysis and Prevention, 7 pp. 281-288.

World Health Organization (2004). *World report on road traffic injury prevention.* World Health Organization.

Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics, 18 pp. 183-190.

Yager, R.R. (1996). Quantifier guided aggregation using OWA operators. International Journal of Intelligent Systems, 11 pp. 49-73.

Yager, R.R. (1997). On the inclusion of importances in OWA aggregations. In The ordered weighted averaging operators. Dordrecht: Kluwer Academic Publishers.

Yager, R.R. and Kacprzyk, J. (1997). The ordered weighted averaging operators: theory and applications. Dordrecht: Kluwer Academic Publishers.

Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network (2005). *2005 environmental sustainability index*. Yale University and Columbia University.

Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network (2006). *Pilot 2006 environmental performance index*. Yale University and Columbia University.

Zador, P.L. (1991). Alcohol related relative risk of fatal driver injuries in relation to driver age and sex. Journal of Studies on Alcohol, 52 pp. 302-310.

Zein, S.R. and Navin, F.P. (2003). *Improving traffic safety: The C3-R3 systems approach*. Presented at the 82nd annual meeting of the Transportation Research Board, Washington D.C.

Zietsman, J., Knowles, W.E., Ramani, T.L., Lee, J.S. and Bochner, B.S. (2008). *Sustainability enhancement tool for state DOTs using performance measurement*. Presented at the 87th annual meeting of the Transportation Research Board, Washington D.C.

Appendix

Appendix I: Illustration of an indicator card

1. Name: Percentage of persons wearing a seat belt in the front seats of a car or van

1.1) Domain: protective systems.

1.2) Description and context: the seat belt wearing rate of drivers and occupants, over 11 years of age, in front of a car or van is determined. The indicator value is the share of persons wearing their seat belt in the total number of persons examined.

1.3) Measurement unit: percentage.

2. Position in the general framework

Seat belt wearing affects the degree of injury in an accident (e.g., World Health Organization, 2004) and is an injury risk factor. The effectiveness of seat belts depends on a number of factors such as the impact speed, the type of collision and the position of the occupant (Elvik and Vaa, 2004). In addition, risk compensation could occur implying that the driver (or occupant) accepts a higher level of risk (e.g., a higher speed) because he feels more safe (see e.g., Gaudry and Lassarre, 2000).

3. Policy relevance

3.1) Applicability or relevance: seat belt wearing in front seats is compulsory in almost all countries. Measuring the wearing rate in practice is of relevance.

3.2) Link with other indicators: a positive correlation with the wearing rate in the rear seats of a car or a van can be expected, as well as with the child's seat use. Furthermore, there may be a link with indicators from other risk domains, for example the age distribution of the car fleet.

3.3) Objective and values to aim at: the Federal Commission for road safety in Belgium instructs to increase the current seat belt objectives in order to obtain a seat belt rate similar to that of the best performing countries in Europe by 2010 (Federale Commissie voor de Verkeersveiligheid, 2007). The target figure for Belgium by 2010 is 95%. By means of various measures – such as campaigns focusing on the favourable effects or a higher level of enforcement – a lot of progress is possible.

3.4) Regulation: for Belgium, article 35 in the traffic regulations (www.wegcode.be) obligates the driver and passengers of cars and other motor vehicles participating in traffic to wear seat belts on the seats equipped with them. More specifically, seat belt wearing in front seats is compulsory since June 1975.

4. Methodological description

4.1) Measurement method: by means of observation the seat belt use of all persons in front of a car or van passing by the observation point will be assessed. In advance, the total number of vehicles to check needs to be determined as well as the locations and times of observation. It is essential that these moments and locations are chosen randomly in order to obtain results that are representative for the entire population.

4.2) Data needed, collection frequency, source: the indicator value can be computed in case two figures are available, being the total number of observed persons and the number of persons wearing their seat belt. Under the authority of the Belgian road safety institute observational seat belt wearing studies are carried out in Belgium every year during May since 2003. Drivers and passengers in the front seats of passenger cars are observed at 150 different locations selected according to region, speed limit and points in time (on a weekday at rush hour, on a weekday outside of rush hour, on a weeknight and during the weekend both day and night) (Godart, 2006).

4.3) Limitations: at locations characterized by high speeds and large intensities, it is impossible to assess the wearing rate of all persons passing by. Clear agreements are needed then. Furthermore, in Belgium about 25,000 passenger cars are currently observed (Godart, 2006) whereas the Federal Commission for road safety instructs to yearly check the occupants of at least 1,800,000 cars (i.e., one vehicle out of three) (Federale Commissie voor de Verkeersveiligheid, 2007).

4.4) Alternatives: applying the same measurement method in several time periods allows studying the evolution over time. A comparison with other countries is also relevant. In accordance with the European recommendations every year at least three integrated control actions need to be organised, each lasting two weeks (Federale Commissie voor de Verkeersveiligheid, 2007). Moreover, the guidelines for seat belt wearing measurement specified in the manual of SafetyNet (Hakkert and Gitelman, 2007) could be helpful.

5. Assessment of the Safety Performance Indicator

5.1) Relevant/valid: the indicator can be associated with a specific performance target and is action-oriented. Moreover, the indicator provides a good picture of the protective systems domain.

5.2) Measurable: the indicator is computed as the ratio of two figures and thus quantifiable.

5.3) Understandable: the indicator has a clear interpretation and is defined in an understandable way.

5.4) Available data: assessing the wearing rates of persons in front of a car or van is common in a number of countries implying the availability of an indicator value for a large set of countries. Moreover, the indicator can be updated on a regular basis.

5.5) Reliable: the indicator values for Belgium are gathered using a scientific measurement method. However, the representativeness of the data could still be enhanced. Moreover, at this point, each country is collecting some data of which the reliability is difficult to assess.

5.6) Comparable/coherent: more or less the same definition is used for this indicator in most countries and over time. Therefore, some comparison is possible.

5.7) Specific: the indicator focuses on persons (older than 11 years) in front of a car or van and is characterized by a useful level of detail. A further division in specific road types could be valuable as well.

5.8) Sensitive: the indicator is sensitive enough; another indicator value implies a change in risk.

6. Evolution of the Safety Performance Indicator

The seat belt wearing rate of drivers and passengers in front of cars in Belgium from 2003 onwards is shown in the graph below. Detailed wearing rates are also available, for example per speed limit or region but here the overall figures are presented. Figure 30 shows an increase in the seat belt wearing rate in front of cars in Belgium over the years 2003-2006. In case the target figures are met in the future, the percentage of persons wearing their seat belt in front of a car will approximate the level of 95% by 2010.

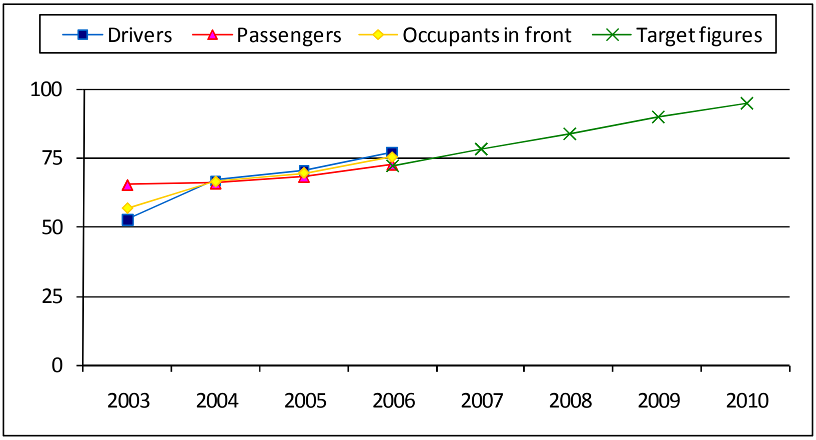


Figure 30: Seat belt wearing rate in front seats of cars in Belgium (2003-2006) and target figures until 2010 (Federale Commissie voor de Verkeersveiligheid, 2007)

7. Comparison between subjects

In Finland, about 60,000 country-wide seat belt observations were performed in 2002 resulting in a frontal seat belt wearing rate of more than 80% inside urban areas and 93% in rural areas (Luukkanen, 2003). It was found that the wearing rate of vans compared to cars was 20 to 30 percent lower.

The European Transport Safety Council (2007) published – front aggregated, front driver, front passenger and rear seats – seat belt wearing rates for a large number of European countries. In 2005, Belgium had a frontal seat belt wearing rate of 71%. Countries performing very well on this indicator (i.e., ≥90%) are France, Germany, the Netherlands, Norway, Sweden and United Kingdom. The average percentage of persons wearing a seat belt in front of a car or a van for the 25 EU member states in 2005 is estimated at 86% (Vis and Van Gent, 2007).

8. Positive and negative aspects

8.1) Strengths: for the risk domain 'protective systems' the seat belt wearing rate in front of cars and vans is a principal indicator. Within this domain, it is the most common indicator for which data are available in most countries for some time, thereby providing insight into relative performance and changes over time. Furthermore, the indicator scores well on the criteria under point 5 (i.e., relevant, measurable, understandable, etc).

8.2) Weaknesses: for this indicator, occupants in the front seats of cars and vans are considered. As there might be a difference in safety behaviour between the two groups, it is important that both vehicle classes are incorporated when comparing the wearing rates across countries. The same holds for the persons checked (i.e., results differ in case both drivers and passengers in front are considered or only one class of occupants) and the locations considered (e.g., speed limit). Moreover, to avoid biased results measurements are needed at regular intervals, thereby testing a large share of vehicles. Guidelines are a prerequisite for a correct, comparable and representative measuring of wearing rates. In case all member states are willing to apply the guidelines developed on the European level within the SafetyNet project (Hakkert and Gitelman, 2007) more reliable indicator data will be obtained enabling better comparisons.

9. Conclusion

Based on the information above, we can conclude that the percentage of persons wearing a seat belt in the front seats of a car or a van is an appropriate indicator for monitoring the protective systems domain.

10. References

Belgian Traffic Regulations. Retrieved January, 30th, 2006, from www.wegcode.be.

Elvik, R. and Vaa, T. (2004). *Handbook of road safety measures*. Oxford: Elsevier Ltd.

European Transport Safety Council (2007). *PIN Flash 4 on seat belt use*. Retrieved March, 5th, 2007, from http://www.etsc.be/PIN-publications.php.

Federale Commissie voor de Verkeersveiligheid (2007). *Staten-Generaal van de Verkeersveiligheid*.

Gaudry, M. and Lassarre, S. (2000). *Structural road accident model: The international DRAG family*. Oxford: Elsevier Ltd.

Godart, B. (2006). Belgen steeds meer gehecht aan de gordel. *Via Secura*, *73*, 8-10.

Hakkert, A. S. and Gitelman, V., (Eds) (2007). *Road safety performance indicators: Manual*. (D3.8 of the EU FP6 project SafetyNet).

Luukkanen, L. (2003). *Safety management system and transport safety performance indicators in Finland*. Likkenneturva.

Vis, M. A. and Van Gent, A. L., (Eds) (2007). *Road safety performance indicators: Country comparisons*. (D3.7a of the EU FP6 project SafetyNet).

World Health Organization (2004). *World report on road traffic injury prevention.* World Health Organization.

Appendix II: Evaluation of possible indicators

Based on literature, eight selection criteria were identified. Next, each possible indicator for a particular risk domain will be evaluated on these criteria. Meeting the criterion is indicated by a '+' sign, a '-' sign refers to the opposite, whereas '0' is the code for a score in between. Other possibilities are '0/+'; '-/0'; and '/' where the last symbol is used for rating reliability; in case no data are available, the degree of reliability is not applicable and not indicated (as a result only seven criteria are used then). To bring transparency in the subjective rating approach, some clarifications are given below the table for the first domain (alcohol and drugs) explaining why a specific code has been assigned. Although the evaluation of all possible indicators is quite time-consuming, it is a necessary exercise and has important implications.

From the evaluation, a set of best needed indicators and a set of best available indicators is deduced. Best needed indicators are ideal indicators with one or a few major shortcomings such as no data available, no reliable data or no comparable/coherent data. Therefore, only five criteria are used for computing the score for the best needed indicator (relevant-measurable-understandable-specific-sensitive), while all eight criteria are considered in the score for the best available indicator. Moreover, the signs attached to each cell in the table are quantified as follows: '+'=1; '0/+'=0.5; '0'=0; '-/0'=-0.5; '-'=-1. The indicator(s) with the highest best needed respectively best available score within each risk domain will appear in the final set of best needed and best available road safety performance indicators.

POSSIBLE ALCOHOL AND DRUGS INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
% of tested road users with alcohol concentration above the legal limit	+	+	+	0	0	-[5]	+	0[7]	4	3/8
% of tested road users with drugs concentration above the legal limit	+	+	+	-	/	-	-/0	0	2.5	0.5/7
% of surveyed (car) drivers disrespecting the alcohol limit	+	+	+	+	0	0[6]	+	0	4	**5/8**
% of road user population impaired by alcohol or drugs	+	+	+	-	/	0	0/+	+	**4.5**	3.5/7
% of road user population impaired by alcohol and drugs	+	+	+	-	/	0	0/+	0/+	4	3/7
% of fatalities resulting from accidents with ≥1 driver impaired by alcohol	0[1]	+	0/+	0/+[4]	0	0/+	+	-/0[8]	2	3/8
% of fatalities resulting from accidents with ≥1 driver impaired by drugs	0	+	0/+	-	/	0	-/0	-/0	0.5	≤0
% of fatal accidents in which someone was drinking and driving	0	+	+	0/+	0	+	+	-/0	2.5	4/8

% of road users involved in fatal accidents impaired by alcohol or drugs	0	+	+	0/+	0	+	0/+	-/0	2	3.5/8
subjective risk of getting caught while drinking and driving	0[(2)]	0	0/+	0/+	0	0	0	0/+	1	1.5/8
attitude towards drinking and driving	0	[(3)]-	0	0	0	0	0	0/+	≤0	≤0

(1) This is an effect indicator, not an action-oriented performance indicator.

(2) This is a state indicator, not an action-oriented performance indicator.

(3) This is a concept that is difficult to measure and quantify.

(4) In some countries only a small part is tested.

(5) The limits differ between countries and changes in limits are possible over time; at present, the measurement method as well as the sampling design are not consistent.

(6) There might be differences in self-reporting behaviour between countries.

(7) A change in enforcement activity might affect the indicator value.

(8) A change in this indicator value might not be due to a better performance.

POSSIBLE SPEED INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
% of tested drivers exceeding the speed limit on ≠ road types	+	+	+	0	0	-/0	0/+	0	3.5	3/8
% of surveyed (car) drivers exceeding the speed limit on ≠ road types	+	+	+	0/+	0	0	0/+	0	3.5	**4/8**
average daytime speed per road type and vehicle type	0	+	+	0	0	-/0	+	+	**4**	3.5/8
average speed per road type and vehicle type at night	0	+	+	-	/	-/0	+	+	**4**	2.5/7
variation in speed per road type and vehicle type	0	+	+	-	/	0	+	+	**4**	3/7
median (or other percentile) of the set of observed speeds divided by the speed limit of the road class	+	+	0	0	0	0	+	0/+	3.5	3.5/8
median of the set of absolute differences between each of the observed speeds in the road class and the median of all observed speeds divided by the median of the set of observed speeds	-	+	-	-	/	0	+	0	≤0	≤0
% of drivers with inappropriate headway on ≠ road types	0	+	+	-	/	-	-	0	1	≤0

POSSIBLE PROTECTIVE SYSTEMS INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
% of persons wearing a seat belt in the front seats of a car or van	+	+	+	+	0	+	0/+	+	**4.5**	**6.5/8**
% of persons wearing a seat belt in the rear seats of a car or van	+	+	+	0/+	-/0	+	0/+	+	**4.5**	5.5/8
% of persons <12 years (correctly) sitting in a child's seat in the front or rear seat of a car	+	+	+	-/0	-/0	+	0/+	+	**4.5**	4.5/8
% of persons wearing a seat belt in the front seats of a bus (>3.5 tons) or a truck	+	+	+	-	/	0	0/+	+	**4.5**	3.5/7
% of persons wearing a seat belt in the passenger seats of a bus	+	+	+	-	/	0	0/+	+	**4.5**	3.5/7
% of cyclists wearing a helmet	0/+	+	+	-	/	-	0/+	+	4	2/7
% of moped riders wearing a helmet	+	+	+	-/0	-/0	0	0/+	+	**4.5**	3.5/8
% of motorcyclists wearing a helmet	+	+	+	-/0	0	+	0/+	+	**4.5**	5/8

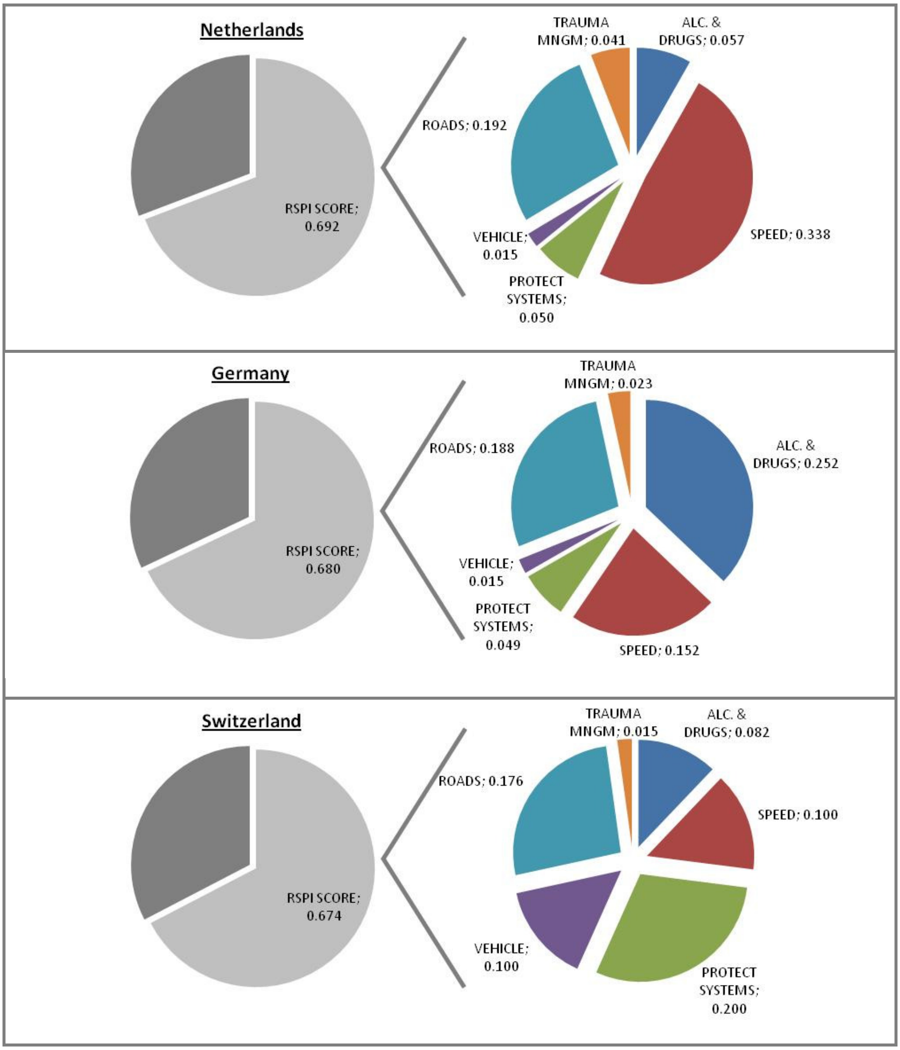
POSSIBLE DAYTIME RUNNING LIGHTS INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
total usage rate of daytime running lights	+	+	+	-/0	-/0	-/0	0	0/+	3.5	2/8
usage rate of daytime running lights per road type and vehicle type	+	+	+	-/0	-/0	-	+	0/+	**4.5**	2.5/8
existence of a law – fully or partially – obligating the use of daytime running lights	0	0/+	+	+	0/+	0/+	0	-/0	1	**3/8**

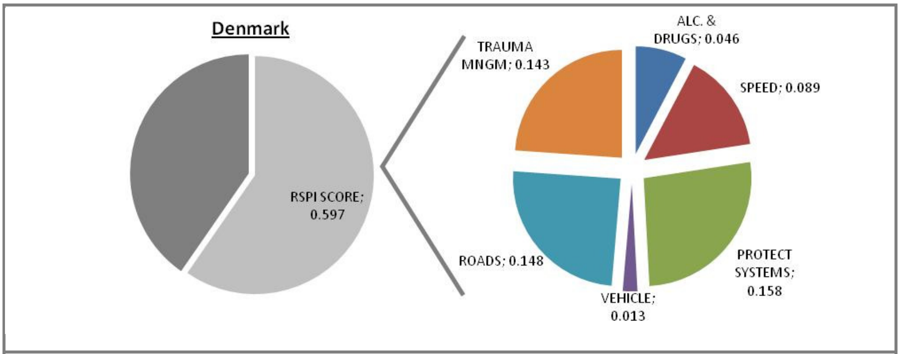
POSSIBLE VEHICLE INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
% of cars failing the official vehicle inspection	0/+	+	+	0	0/+	0/+	0	+	3.5	4.5/8
EuroNCAP score: % of new cars obtaining 0 respectively 1, 2, 3, 4 or 5 stars in total # of new passenger cars	0/+	+	+	0	0/+	0/+	0	+	3.5	4.5/8
age distribution of the vehicle fleet: % of vehicles ≤5 years; between 6-10 years, between 11-15 years and >15 years in the total # of registered vehicles (focus on passenger cars)	0/+	+	+	0/+	0/+	+	+	0/+	**4**	**6/8**
composition of the vehicle fleet: % of cars, vans, buses, trucks and motorcycles in total # of registered vehicles	0/+	+	+	+	0/+	+	0	0	2.5	5/8

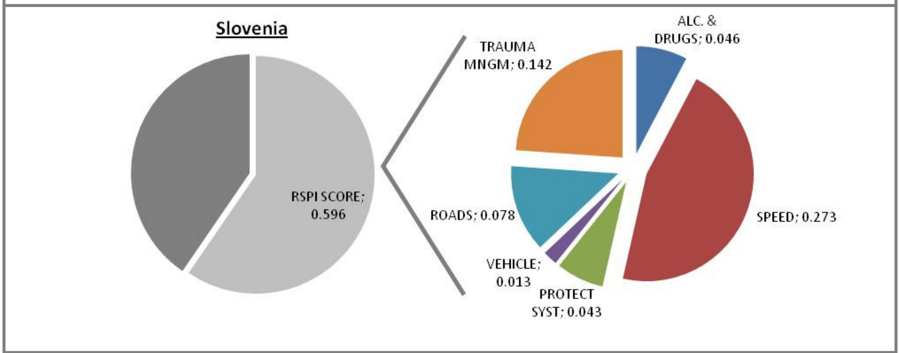
POSSIBLE ROADS INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
share of network length per road type	0	+	+	0/+	+	-/0	+	-	2	3/8
share of intersections per type	0	+	+	-	/	0	+	0	3	2/7
intersection density	0	+	+	0	0	+	-	-	≤0	1/8
network density	0	+	+	+	+	+	-	-	≤0	3/8
motorway density	0	+	+	+	+	+	+	-	2	**5/8**
EuroRAP road protection scores	+	0/+	0	-/0	0	+	0	0/+	2	2.5/8
% of road length with a wide median or median barrier	+	+	0	-	/	0	+	0/+	**3.5**	2.5/7
% of road length with a wide obstacle-free zone or roadside barrier	+	+	0	-	/	0	+	0/+	**3.5**	2.5/7
% of road length with facilities for separation of slow vulnerable traffic and other motorized traffic	+	+	0	-	/	0	0/+	0/+	3	2/7
% of road network satisfying the safety design standard	+	0	0	-	/	-/0	-	0/+	0.5	≤0
expenditure on roads as GDP share	0/+	+	0/+	+	+	0	-	0	1	3/8

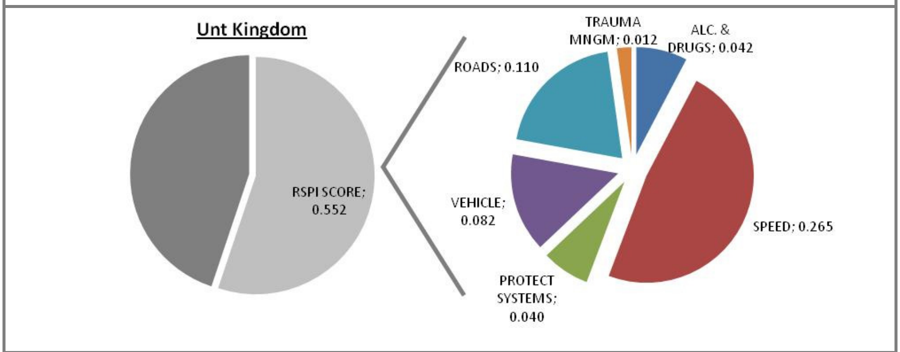
POSSIBLE TRAUMA MANAGEMENT INDICATORS	**Relevant/ valid**	**Measurable**	**Understan-dable**	**Available data**	**Reliable**	**Comparable/ coherent**	**Specific**	**Sensitive**	**Scores best needed (/5)**	**Scores best available**
% of calls to emergency medical services due to a road accident	0/+	+	+	-	/	0	+	0	3.5	2.5/7
average arrival time of emergency medical services at the accident scene	+	+	+	-	/	-/0	+	0	**4**	2.5/7
# of emergency medical services' staff per 10,000 citizens	0	+	+	0	+	0	+	0	3	4/8
# of hospital beds per 10,000 citizens	0	+	+	+	+	0	0	-/0	1.5	3.5/8
average length of stay in the hospital after a road accident	0/+	+	+	-	/	0	+	0	3.5	2.5/7
share of road casualties treated in intensive care units	0/+	+	+	-	/	0/+	+	0	3.5	3/7
share of road casualties who died during hospitalisation	+	+	+	-	/	0/+	+	0	**4**	3.5/7
expenditure on health care as GDP share	0/+	+	+	+	+	0	0	0	2.5	**4.5/8**

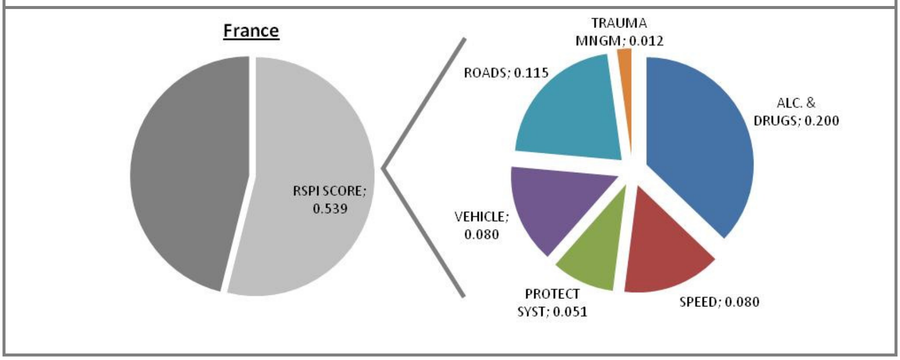
Appendix III: Visualisation of the road safety performance index score and its components per country

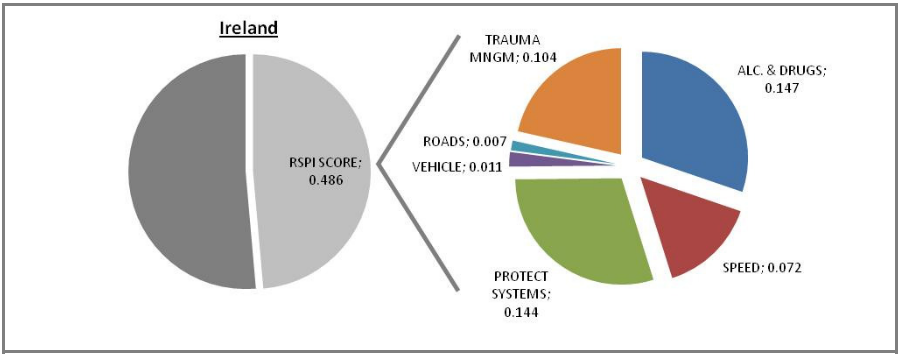


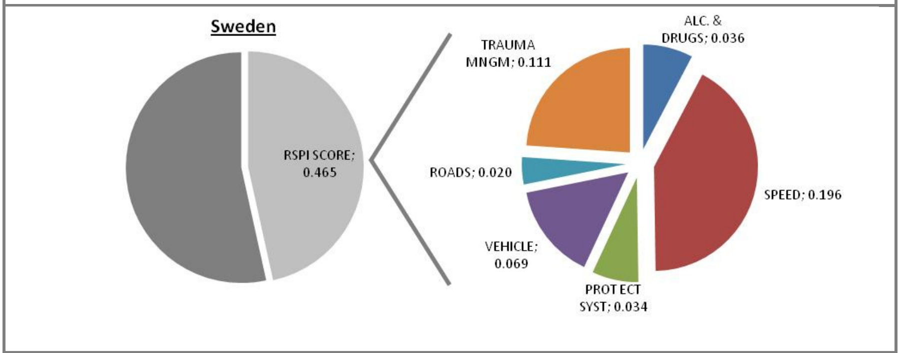

Denmark
RSPI SCORE; 0.597
TRAUMA MNGM; 0.143
ALC. & DRUGS; 0.046
SPEED; 0.089
PROTECT SYSTEMS; 0.158
VEHICLE; 0.013
ROADS; 0.148

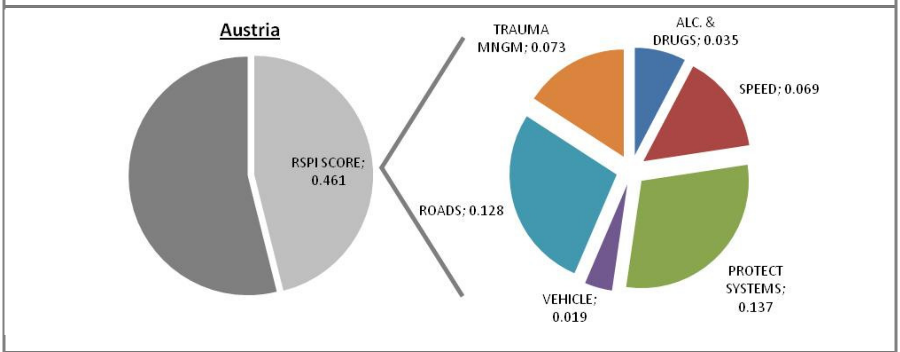

Slovenia
RSPI SCORE; 0.596
TRAUMA MNGM; 0.142
ALC. & DRUGS; 0.046
SPEED; 0.273
PROTECT SYST; 0.043
VEHICLE; 0.013
ROADS; 0.078

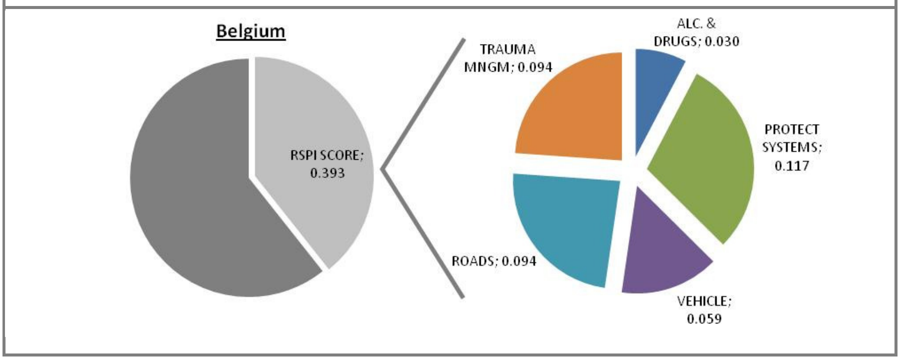

Unt Kingdom
RSPI SCORE; 0.552
TRAUMA MNGM; 0.012
ALC. & DRUGS; 0.042
ROADS; 0.110
SPEED; 0.265
VEHICLE; 0.082
PROTECT SYSTEMS; 0.040

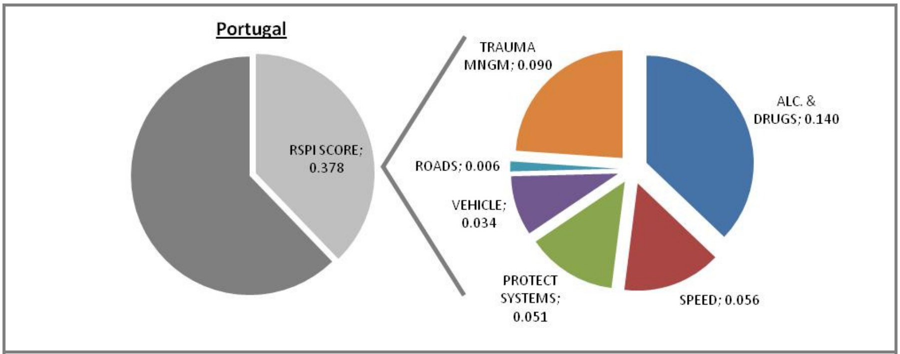

France
RSPI SCORE; 0.539
TRAUMA MNGM; 0.012
ROADS; 0.115
ALC. & DRUGS; 0.200
VEHICLE; 0.080
PROTECT SYST; 0.051
SPEED; 0.080

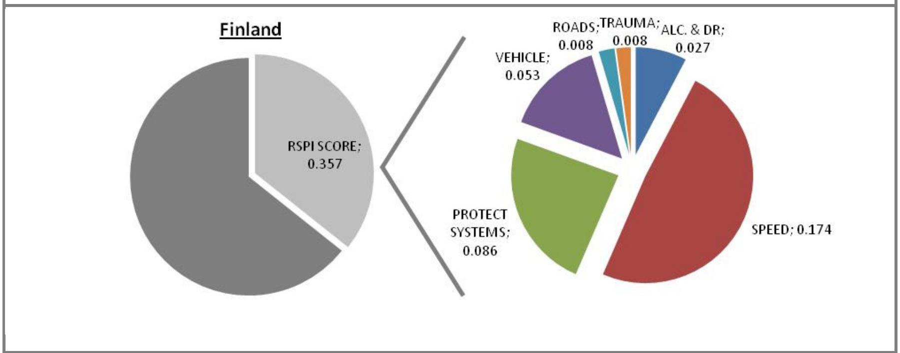

Ireland
RSPI SCORE; 0.486
TRAUMA MNGM; 0.104
ALC. & DRUGS; 0.147
ROADS; 0.007
VEHICLE; 0.011
SPEED; 0.072
PROTECT SYSTEMS; 0.144

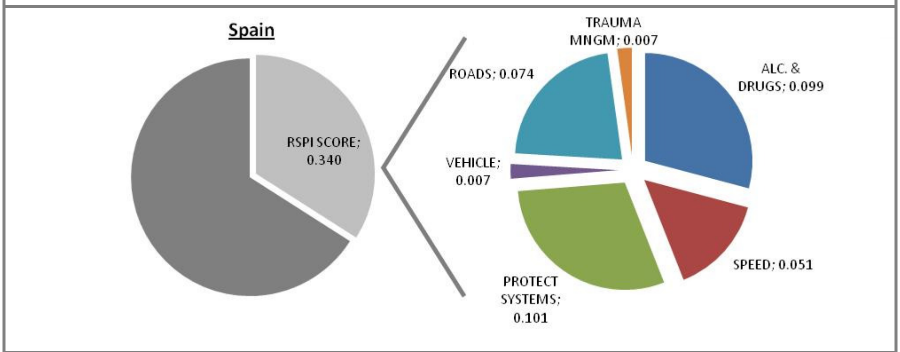

Sweden
RSPI SCORE; 0.465
TRAUMA MNGM; 0.111
ALC. & DRUGS; 0.036
ROADS; 0.020
SPEED; 0.196
VEHICLE; 0.069
PROT ECT SYST; 0.034

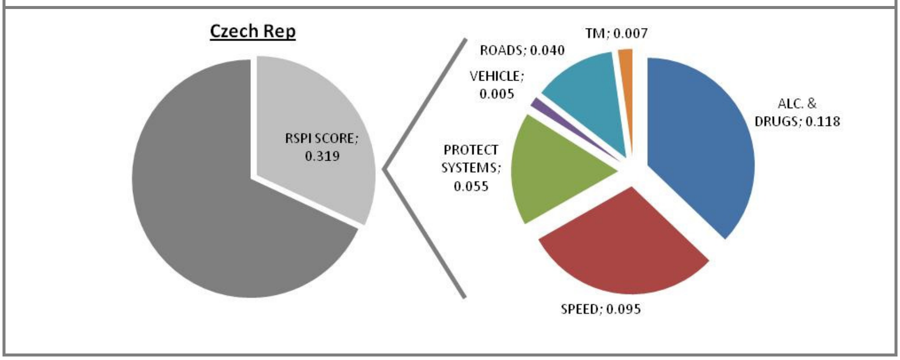

Austria
RSPI SCORE; 0.461
TRAUMA MNGM; 0.073
ALC. & DRUGS; 0.035
SPEED; 0.069
ROADS; 0.128
PROTECT SYSTEMS; 0.137
VEHICLE; 0.019

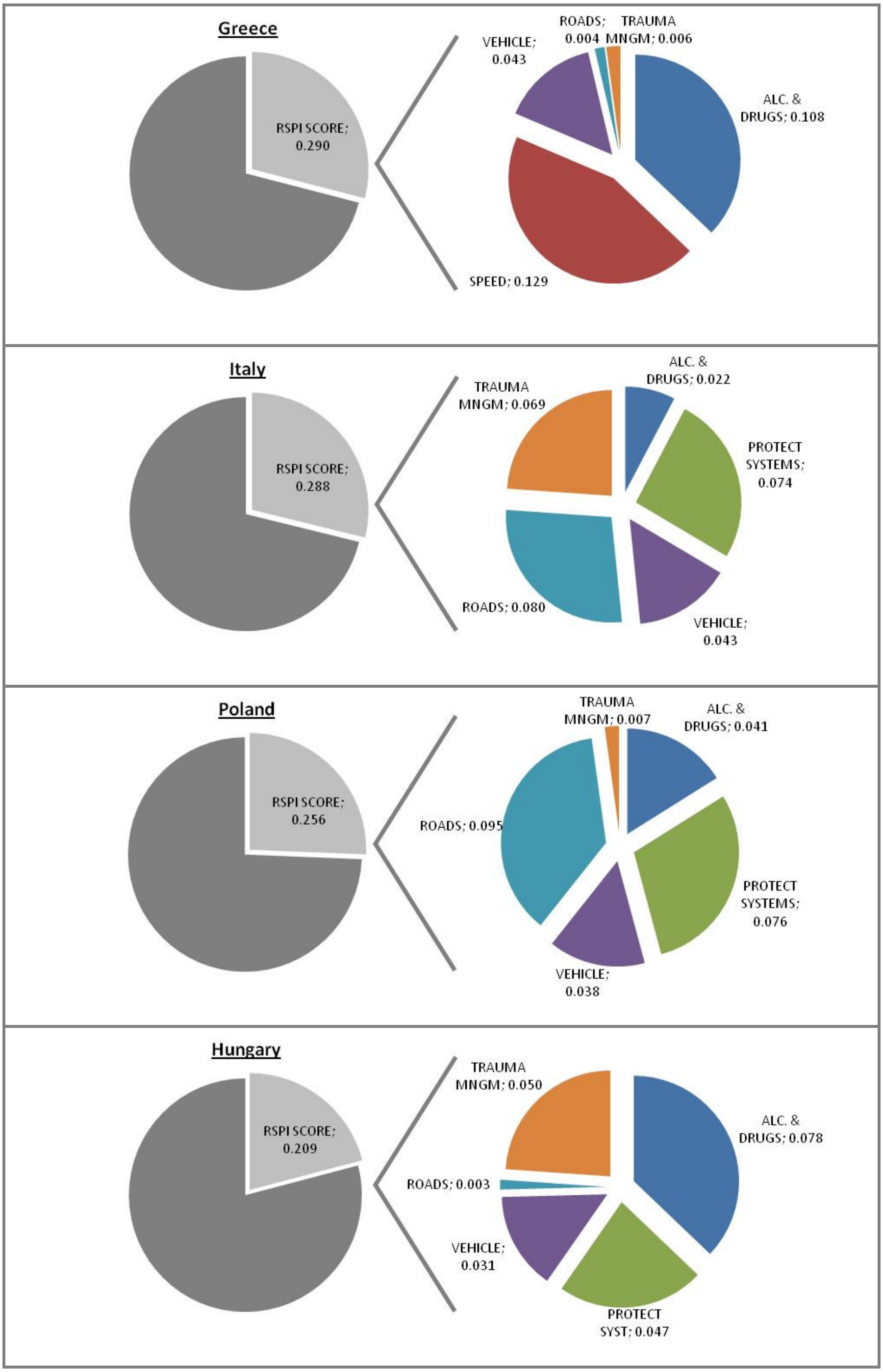

Belgium
RSPI SCORE; 0.393
TRAUMA MNGM; 0.094
ALC. & DRUGS; 0.030
PROTECT SYSTEMS; 0.117
ROADS; 0.094
VEHICLE; 0.059

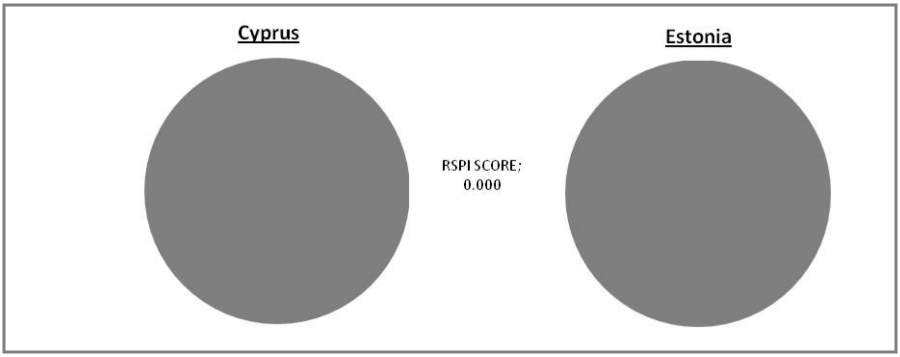

Portugal
RSPI SCORE; 0.378
TRAUMA MNGM; 0.090
ALC. & DRUGS; 0.140
ROADS; 0.006
VEHICLE; 0.034
PROTECT SYSTEMS; 0.051
SPEED; 0.056


Finland
RSPI SCORE; 0.357
ROADS; 0.008
TRAUMA; 0.008
ALC. & DR; 0.027
VEHICLE; 0.053
PROTECT SYSTEMS; 0.086
SPEED; 0.174


Spain
RSPI SCORE; 0.340
TRAUMA MNGM; 0.007
ROADS; 0.074
ALC. & DRUGS; 0.099
VEHICLE; 0.007
SPEED; 0.051
PROTECT SYSTEMS; 0.101


Czech Rep
RSPI SCORE; 0.319
TM; 0.007
ROADS; 0.040
VEHICLE; 0.005
ALC. & DRUGS; 0.118
PROTECT SYSTEMS; 0.055
SPEED; 0.095


Greece
RSPI SCORE; 0.290
ROADS; 0.004
TRAUMA MNGM; 0.006
VEHICLE; 0.043
ALC. & DRUGS; 0.108
SPEED; 0.129
Italy
RSPI SCORE; 0.288
ALC. & DRUGS; 0.022
TRAUMA MNGM; 0.069
PROTECT SYSTEMS; 0.074
ROADS; 0.080
VEHICLE; 0.043
Poland
RSPI SCORE; 0.256
TRAUMA MNGM; 0.007
ALC. & DRUGS; 0.041
ROADS; 0.095
PROTECT SYSTEMS; 0.076
VEHICLE; 0.038
Hungary
RSPI SCORE; 0.209
TRAUMA MNGM; 0.050
ALC. & DRUGS; 0.078
ROADS; 0.003
VEHICLE; 0.031
PROTECT SYST; 0.047


Cyprus
Estonia
RSPI SCORE;
0.000