Safety Science 48 (2010) 1203–1211

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Benchmarking road safety performances of countries

Fred Wegman *, Siem Oppe

SWOV Institute for Road Safety Research, Delft University of Technology, Faculty Civil Engineering & Geosciences, The Netherlands

ARTICLE INFO

Article history:
Received 14 September 2009
Received in revised form 31 December 2009
Accepted 4 February 2010

Keywords:
Road safety
Performance indicators
International comparison
Model
Policy evaluation

ABSTRACT

In order to obtain political interest in road safety problems and to learn from other countries' 'good practices', it is often helpful to compare one's own safety situation with that of other countries. In a number of projects tools have been developed for such comparisons. These tools range from simple ratings of countries on their safety outcomes, such as the annual number of fatalities per capita or per kilometre driven by (motor)vehicles to more comprehensive comparisons.

These comparisons not only show differences in safety between countries, but to a certain extent also explain such differences in terms of their safety background and measures taken. Finally, tools have been defined to support road safety policy makers in developing possible safety measures or actions. Procedures for such complex safety comparisons have been developed and tested in several so-called SUNflower studies.

This promising approach can be further developed into standard procedures for safety comparisons between all countries in the European Union, and other countries worldwide. This paper wishes to outline the development of such standards for the benchmarking of road safety and safety trends as well as procedures for quantifying safety performances of countries.

Starting point of this conceptual framework is the so-called SUNflower-pyramid in which three types of indicators are distinguished. The first one of these, the road safety performance indicator, is called an outcome indicator and is based on the number of killed and injured road users. The second indicator type indicates the quality of the implementation of road safety policies: the implementation performance indicators. The third type of indicator indicates the quality of response in policy documents to improve road safety (policy performance indicator). The three types of indicators are embedded in a policy context: the structure and culture of a country, which are considered as background variables.

This paper sets out to describe the framework for the development of a comprehensive set of indicators to benchmark road safety performances of countries or of sub-national jurisdictions. The paper also discusses the advantages and disadvantages of combining such indicators and if combined, how to aggregate how different indicators in one composite performance index. It is argued to group countries in different classes with more or less comparable countries. Different procedures are used for this grouping. The results are promising and it is recommended to work with classes of countries.

1. Introduction

It is important for countries to compare their safety performances with those of other countries. A first motivation for comparison is to know how the overall safety situation in the most recent years compares with that in other countries. Based on this we expect to learn from other countries and to identify advanced policies in use abroad in order to apply them in one's own country. Sometimes the comparisons are expressed in terms of rankings. In order to do so, it is necessary to define safety. Safety is often defined in terms of mortality rates: fatalities per head of the population. Mortality rates are used primarily to rank road safety or traffic risk to other risks, such as mortality due to diseases, during labour accidents, or accidents in and around the house. For the comparison of traffic risks this has the disadvantage that the level of motorization is not taken into account. Therefore, another indicator is commonly used as a criterion for traffic safety: fatality risk, defined as the number of fatalities per motor vehicle kilometre. For those countries in which the motor vehicle kilometres are not available, the fatality rate – defined as the number of fatalities per motor vehicle – will be used instead.

Not only the recent safety situation is of interest, but also the safety development over time: has the country's safety been increasing or decreasing over time? Therefore, trend analyses should be carried out to enable comparisons between countries over time.

* Corresponding author. Tel.: +31 653693816.
E-mail address: fred.wegman@swov.nl (F. Wegman).

0925-7535/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2010.02.003

For all types of comparison the most important question is: which country do we want to compare ourselves with? This question is not easily answered. The answer depends on the purpose of the comparison. If only a simple ranking of countries in a certain year is required, then the fatality risk indicator seems to provide sufficient information. However, even then it is not fair to carry out a direct comparison between all countries. Some countries have a more difficult task to fulfil than others and a correction for such a handicap should be applied. However, it is not easy to define and measure such a handicap. Basically this statement on 'handicaps' goes back to two observations. First of all, weaknesses and opportunities of different countries should be identified in the structure and culture layer. To mention two examples only: a strong leadership in a country to implement effective road safety measures, or a willingness in a society to accept freedom-restricting legislation and enforcement to improve road safety such as lowering speed limits and drinking at driving limits. Another observation is that if a country has taken all well-known, easy to implement road safety measures, new approaches have to develop to make further progress. This may result in a stage of diminishing marginal return of investments, and consequently, in less support from society and politics.

If a comparison over time needs to be carried out, then the situation is even more complicated. In that case there is not a single indicator that unambiguously ranks countries. It is not easy to define 'the-best-in-class' this way. Nor is a comparison of a large number of trends easy to make. Comparisons between a smaller range of countries with similar traffic systems or safety levels, or with a more general common background seem to be more promising. Of course countries can learn from measures taken in all other countries. But to formulate targets or plans it is more realistic to compare with countries in the same situation, and/or with the same economical, historical and geographical background, and/or the same level of motorization and safety development as resulted from the SUNflower+6 study (Wegman et al., 2005).

To address this problem of meaningful comparisons in the field of road safety, the SUNflower approach was developed. SUNflower is the acronym of a series of projects. These studies started with a comparison of road safety developments in Sweden, the United Kingdom and the Netherlands (Koornstra et al., 2002). Later this approach was extended to include six other European countries: SUNflower+6 in which three groups of three countries compared themselves using the same methodology (Wegman et al., 2005). These countries include the traditional SUN countries, three countries in the South of Europe: Greece, Portugal, Spain (with special emphasis on Catalonia) and three in Central Europe: the Czech Republic, Hungary and Slovenia. The third and most recent study, SUNflowerNext (Wegman et al., 2008), developed a framework for benchmarking the safety performance of countries and made first attempts to capture this process in a safety performance index.

This paper introduces the SUNflowerNext approach. This approach aims to develop a knowledge-based framework for comprehensive benchmarking of road safety performances and developments of a country or of sub-national jurisdictions. Benchmarking is the search and implementation of best practices. For this benchmarking we introduce different indicators for road safety performances (Section 3). These different indicators will be combined in a composite index and the pros and cons of this approach are discussed (Section 4). Three different types of performance indicators are discussed (Section 5). Benchmarking compares the performance of a country with other so-called 'best-in-class' practices (in other countries). How to form 'classes' is subject of Section 6. The paper concludes with a section on conclusions (Section 7).

2. The SUNflower approach

The SUNflower approach attempts to answer the question what exactly caused road safety to improve in (SUN) countries. When specific beneficial aspects of measures, operational practices, or underlying concepts can be determined, what are the possibilities for transferring these aspects to another SUN country or other countries? A better insight into the relationship between the developments of traffic risks and road safety policies, programmes and measures in these countries might conceivably identify key actors, which could further improve the current road safety practice in each of the SUN countries, and in other countries.

The methodological approach is based on a road safety target hierarchy as shown in Fig. 1 and was adapted from the consultation document on the Road Safety Strategy 2010 of New Zealand (LTSA, 2000). In this approach a fundamental understanding is required of traffic safety processes at different levels in the hierarchy of causes and consequences that lead to casualties and costs for society. The main reference is the model that describes a target hierarchy of 'structure and culture' towards 'social costs' (Koornstra et al., 2002).

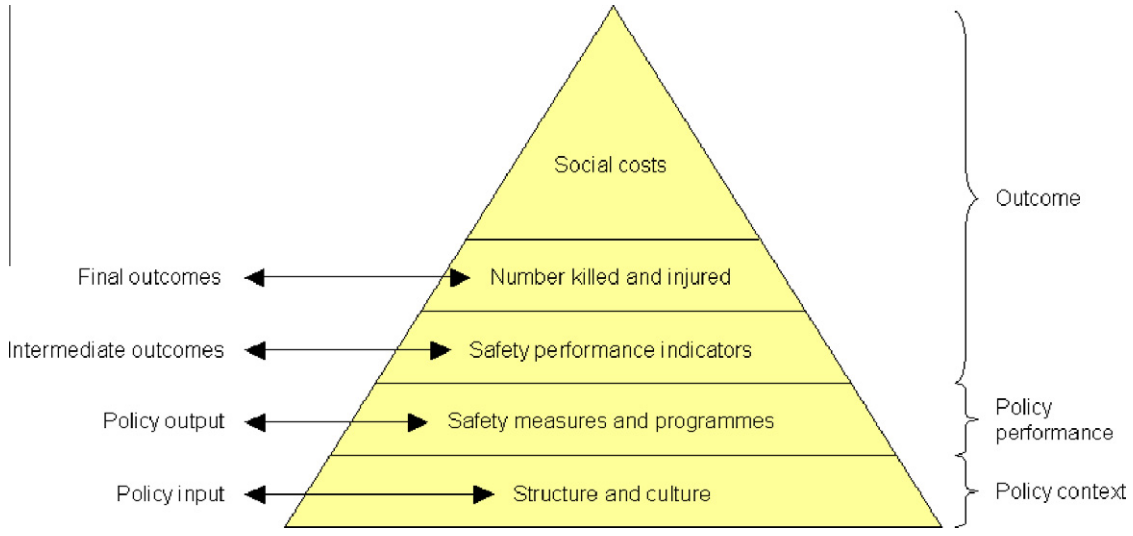

Fig. 1. A target hierarchy for road safety (Koornstra et al., 2002; LTSA, 2000).

In addition to the New Zealand approach we added the 'structure and culture' layer. It was argued that such a basic layer was necessary as a policy context for understanding (impacts of) road safety policy. The component 'structure' addresses two dimensions: the physical structure and the operational (functional) structure. The 'culture' element concentrates on how a society and its citizens perceive the road safety problem compared with, for example, the role of (motorized) traffic in our society and its contribution to economic growth, welfare, environmental consequences, etc. Furthermore, questions are relevant on how responsibilities are defined for individuals and the government. We can also use the term 'safety culture' here (AAA, 2007).

The vertical dimension is formed by the different levels of the pyramid. The traditional way to describe the safety performance (outcomes of the system) is by using the number of killed and injured, indicated in the pyramid by final outcomes. Going up one layer, the top of the pyramid is formed by the social costs, which can be related to the number of casualties and damage. Going down, the level safety measures and programmes reflects the policy performance, or the extent to which policy makers organize safety policies. Policy output should lead to an increased safety quality of the traffic system, which is reflected by better operational conditions for every day traffic (e.g. quality of roads, vehicles, behaviour). The indicators at this level are called safety performance indicators (SPI), and are the intermediate outcomes between the policy output and the number of casualties. SPI's can predict safety levels before accidents have happened, assuming that causal relationships are known. At the bottom level, the structure and culture of a country describe the policy context such as public attitudes towards risk and safety, the organisation of a country, and its history and cultural background. These items should always be taken into account when customizing one country's measures to another country.

In order to understand a country's road safety performance, one can move through the pyramid in both directions: bottom-up or top-down. The SUNflower approach uses case studies for this purpose. For instance, from a sociological point of view, one can first describe public attitudes towards drinking or speeding (structure and culture) and climb the pyramid to identify measures (such as legal limits and enforcement activity) and consequently understand the extent of misbehaviour, and the related casualties and costs. From a cost-effectiveness point of view, one can take the opposite direction, by identifying which problems are responsible for the highest costs, tracking those problems down to their origin, and solving them in a cost effective way. It is not necessary to start only at the bottom or only at the top level, for instance when tracking the effects of established safety measures. Moreover, some mechanisms are not bound by the sequence of pyramid levels. For instance, a change of casualty numbers or the occurrence of a severe accident with a lot of publicity may affect public attitudes directly.

Each level of the pyramid contains several problems, events, or safety topics. The performance of a country with respect to these problems is a reflection of its road safety performance. These problems can be disentangled into the components of the traffic system which constitute the structure of each level, called the second (horizontal) dimension. If we analyse a case study we select part of the horizontal dimension of each of the layers of the pyramid. For this purpose the traditional components vehicle, road and road user aspects could be used (Haddon, 1968). These can be subdivided into vehicle types, road types, user groups, age groups and typical behavioural aspects. A differentiation in regions within a country, seasons within a year or types of casualties can also be introduced here. The actual subdivision may be different for each level, but overlap and interaction is aimed for as much as possible. Subsequently, developments of factors in both the horizontal dimension and the vertical dimension can be tracked over time, the third dimension.

The SUNflower approach used comparisons of three (European) countries. When comparing only two countries it is never easy to interpret observed differences between the two (is one high, or the other one low?), but when three countries are compared, the result is always that two are somewhat closer to each other and the third one is further away. So both in the original SUNflower study (Koornstra et al., 2002) and in the SUNflower+6 study (Wegman et al., 2005) comparing three countries turned out to be valuable for generating potential explanations for observed differences. Furthermore, it turned out that comparing 'comparable' countries has a certain advantage. Although this was not a formal part of both studies, it became apparent that the three clusters in SUNflower+6 (SUN, South, Central) were more involved and interested in each other. On the other hand it is not a proven fact, nor a hard and fast rule, that comparing three countries is the only or the best approach.

In working with this 'SUNflower-approach' it was felt a necessity to condense the vast amount of information in indicators or indices in a concise and comprehensive way. This could be done in two steps. First of all, indicators for road safety performances must be identified (see Section 3). And as a second step, indicators must be brought together in one index (see Section 4).

3. Indicators for road safety performances

Three main functions of indicators have been established by Adriaanse (1993) in his efforts to build indicators to be used in environmental policies: simplification, quantification and communication. This implies that defining indicators should be directed by the intentions which must be satisfied when using the indicators. Basically, by using indicators we try to capture complex phenomena in relatively simple terms and in doing this we run the risk of losing relevant information or insights. Nevertheless, according to Adriaanse, indicators generally use simplification to make complex phenomena quantifiable in such a manner that communication is either enabled or promoted. Furthermore, these indicators can be used to compare countries, to rank them and to benchmark them.

Wegman et al. (2008) proposed to distinguish three types of indicators and in doing so try to condense the vast amount of information. The first type of indicator captures the quality of road safety in a country and has been named road safety performance indicator. The terms 'outcome indicator' and 'product indicator' are also used. In SUNflower we distinguish final outcomes (numbers of killed and injured), intermediate outcomes (such as the safety performance indicator), and social costs. In this article we use the top three layers of the pyramid (see Fig. 1) to indicate a country's road safety performance.

For a meaningful comparison of countries, numbers of people killed or injured are typically 'normalized', which results in fatality rates, the fatalities per, for example, inhabitant, vehicle, or kilometre travelled. In addition, more vulnerable groups of road users like pedestrians, cyclists, and motorized two-wheelers may be specifically considered. Based on the number of people killed and injured and the consequences of road accidents, it is possible to express the socio-economic burden imposed on societies by road crashes. These costs enable a comparison of the consequences of road crashes with other threats to public health (Peden et al., 2004) or with other investment priorities in a country or jurisdiction (Elvik and Veisten, 2005).

The second type of indicator regards the quality of the implementation of road safety policies and has therefore been named Implementation performance indicator. For this implementation

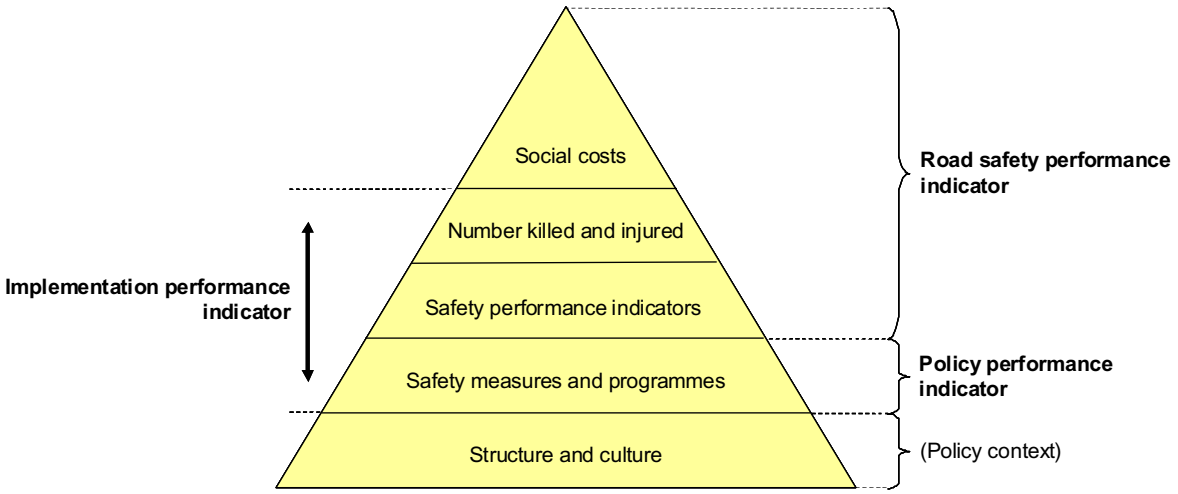

Fig. 2. Three types of performance indicators captured in the SUNflower-pyramid.

performance indicator the term 'process indicator' can also be used. Basically, this indicator follows a vertical line in the pyramid linking 'safety measures and programmes', safety performance indicators, and the numbers of people killed and injured. This indicator will also include elements of the 'structure and culture layer'.

The third type of indicator regards the quality of response in policy documents to improve road safety and has been named policy performance indicator. It has two components: the quality of conditions (strategies, programmes, resources, coordination, institutional settings, etc.) and the quality of action plans and individual (counter)measures in the perspective of ambitions of countries as expressed in road safety targets.

Policy performance deals with the quality of road safety strategy, more specifically with the quality of road safety plans, and with the conditions introduced for the actual implementation of road safety measures and programmes, e.g. institutional arrangements, budget, application of evidence-based knowledge, sound analysis and diagnosis of road safety problems, vertical cooperation between different tiers of government, et cetera. Reports by OECD (2002), ETSC (2006), Bliss and Breen (2009) summarize the demands for the effective development and implementation of national road safety policies.

Hence, three types of indicator, road safety performance indicator, Implementation performance indicator and policy performances indicator are included in the SUNflower-pyramid (Fig. 2).

4. A composite index for benchmarking road safety performances

Essentially, the purpose of (international) cooperation is to learn from each other. This learning should be targeted at a better understanding of the subject involved, in our case road safety. The learning includes subjects such as monitoring and explaining road safety developments, and gaining insights in the impacts of interventions (in the causal relationships between interventions and impacts on road safety, in the active ingredients of interventions and in the dose–response relationships) as a basis to speed up road safety improvements in one's own country or jurisdiction.

Benchmarking is a process in which countries or sub-national jurisdictions evaluate various aspects of their performance in relation to other, and so-called 'best-in-class' practices. The benchmark results provide countries or jurisdictions with information from other international parties that can be used as a basis for developing measures and programmes to increase their own performance. The concept of comparing performances and, one step further, that of benchmarking performances seems to be an appropriate approach for road safety. Benchmarking was originally developed and used in the private sector to compare the performances of individual companies as a tool for improving their operations (Camp, 1989). Benchmarking sets out to provide an objective way of measuring performances against a meaningful reference (in the private sector: the competitor). This meaningful reference is sometimes described as the performance of the 'best-in-class'. This comparison is usually made with the aim of increasing some aspects of one's own performance, in other words to learn from each other.

Benchmarking of road safety performances consists of the following core activities: identifying the key components of road safety performance, identifying whom to compare with (other countries/jurisdictions and 'best-in-class'), constructing indicators for meaningful comparisons, determining and understanding gaps in performances, and, finally, establishing future attainable performances. It is attractive to speak about a benchmark cycle and to carry out benchmarking at regular intervals, to monitor progress made and to evaluate the results of interventions.

Two important tasks can be identified for this process:

- Defining the key components of a road safety performance and investigating if and how these key components can be brought together in a composite index, a road safety performance index.
- Finding a meaningful 'reference' (best-in-class) and defining procedures for identifying such a meaningful reference.

Examples of such indices are known in other domains such as the human development index, which reflects life expectancy, education level and living standards in each country, and is used by the United Nations for the estimation of progress and annual country comparisons, the environmental sustainability index which is used by the world economic forum, the overall health system index used by the world health organisation (WHO), and others (Nardo et al., 2005).

In comparing the safety achievements of countries there is a need to reduce the dimensions of the problem so as to be able to work with a composite index that can express all the relevant components in a concise and comprehensive way. For several reasons, it is attractive to combine all information in one indicator, a so-called composite index. However, advantages and disadvantages must be balanced.

The main pros and cons of using composite indices are adapted from Saisana and Tarantola (2002) by OECD/JRC (2008).

Pros

Composite indices:

- Can summarize complex, multi-dimensional realities with a view to supporting decision-makers.
- Are easier to interpret than a battery of many separate indicators.
- Can assess progress of countries over time.
- Reduce the visible size of a set of indicators without dropping the underlying information base.
- Thus make it possible to include more information within the existing size limit.
- Place issues of a country's performance and progress at the centre of the policy arena.
- Facilitate communication with the general public (i.e. citizens, media, etc.) and promote accountability.
- Enable users to compare complex dimensions effectively.

Cons

Composite indices:

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

Nardo et al. (2005) quote several statements on the pros and cons of using composite indices and these quotations illustrate these pros and cons rather well. To give one example:

"The aggregators believe there are two major reasons that there is value in combining indicators in some manner to produce a bottom line. They believe that such a summary statistic can indeed capture reality and is meaningful, and that stressing the bottom line is extremely useful in garnering media interest and hence the attention of policy makers. The second school, the non-aggregators, believes one should stop once an appropriate set of indicators has been created and not go the further step of producing a composite index. Their key objection to aggregation is what they see as the arbitrary nature of the weighting process by which the variables are combined." (Sharpe, 2004, in Nardo, 2005.)

The OECD (Nardo et al., 2005) prepared a handbook on constructing composite indicators with the aim to provide a guide for constructing and using composite indices for policy makers, academics, the media and other interested parties. The handbook is concerned with indicators which compare and rank country performance in areas such as industrial competitiveness, sustainable development, globalization and innovation. The handbook contains a set of technical guidelines that can help constructors of composite indices to improve the quality of their outputs."

We consider the main advantage of coming to an index the opportunity to raise interests from politicians, policy makers and legislators, interest groups and media. In other policy fields this turns out to be the case. Furthermore we concluded that steps as defined in the OECD handbook are possible to make." (Hermans, 2009; Wegman et al., 2008).

5. Introduction of road safety performance indicators

We distinguish three types of performance indicators for road safety (see Section 3):

- Road safety performance indicators.
- Policy performance indicators.
- Implementation performance indicators.

One can observe quite some activities in defining and collecting data on road safety performance indicators. From a historical perspective, this topic has received a lot of attention and the European Road Safety Observatory, the IRTAD database are valuable sources for these indicators. Both other indicators are relatively less well developed; the more reason to pay extra attention to them.

In recent years, there has been increasing attention for rational decision-making about road safety in many countries in the world. This was also due to the fact that many countries use quantitative targets. This indicates a policy and research based interest in the road safety developments (will we or would not we make the target?), as well as in the factors that can explain those developments (did the implemented policy make a contribution to the developments that were observed?).

5.1. Benchmarking policy performances

Adriaanse (1993) identifies four key factors as crucial for policy performance indicators:

- Quality aspects.
- Sensitivity in time.
- Policy relevance.
- Recognizability and clarity.

The first factor concerns not only the quality of data, but also the methodology used. This methodology must be clearly defined, accurately described, socially and scientifically acceptable and, consequently, easy to reproduce. Furthermore, the indicators should enable composing temporal trends and identifying effects of medium-term and long-term policy interventions. The developed methodology must be derived from the main policy structure, paying attention to major policy themes and target groups, not only for a country as a whole, but also at a sub-national level. Finally, recognizability and clarity require carefully designed indicators with an easy appeal in order for them to be accepted by policy makers and scientists.

Policy performance deals with the quality of a road safety strategy. Wegman (2004) listed the items that are required for the evaluation of policy documents, such as the political support of policy documents, the active support of stakeholders, the precision of the definitions of goals/objectives/targets, the use of valid causal theories on the relations between problems and solutions. At this stage, the assessment of policy documents, resulting in a score, is based on expert opinions. However, it is recommended to develop a standardized methodology for this purpose and to collect data for this indicator at a regular basis.

5.2. Benchmarking implementation performances

Benchmarking implementation performance mainly involves the causal relations between the different layers in the pyramid. Based on international literature on public policies (Wegman, 2004) lists the circumstances that affect the implementation quality of the policy documents and that are useful for monitoring progress:

- Economic/social/political environment.
- Public support.
- Progress of the implementation of policy documents.
- Support of key stakeholders.
- Quality of the 'delivery mechanisms'.

In this context, the World Bank report (Bliss and Breen, 2009) is interesting. This report is based on the six recommendations in the WHO/World Bank report on road safety (Peden et al., 2004). The World Bank report can be regarded as a further elaboration and detailing of the list that was formulated by Wegman (2004).

It has been clearly illustrated above that a good understanding of the effect of a specific road safety measure is only possible if the embedding of such a measure is sufficiently known. This has two major consequences. If it is important to know whether a measure has been implemented well or whether there is room for improvement, it is wise to not only consider the actual measure, but to pay special attention to its embedding. Making explicit advance statements about a measure's expected output increases the possibilities of getting a thorough understanding of the effects of measures. There is also a second consequence: if a country wants to learn from another while the circumstances differ, not only the measure itself, but also its embedding needs to be taken into account. Only then can insight be attained in the possibilities of successfully 'importing' a measure that has been effective elsewhere.

Bliss and Breen (2009), discuss seven management functions of which 'results focus' (the ambition to improve road safety and the means agreed to achieve this ambition) is the most prominent one. The other functions brought forward are coordination, legislation, funding and resource allocation, promotion, monitoring and evaluation, and, finally, research and development and knowledge transfer. Their report rather emphasizes the government's role in increasing road safety and, indeed, there are no examples of substantial and lasting progress without the government having a (very) prominent role. Therefore, it is understandable that the WHO/World Bank report (Peden et al., 2004) argues for a 'lead agency' as a driving force behind the initiatives for and the implementation of road safety policy.

Finally, we will need to address the individual responsibility of citizen and road user, and to answer the question whether it is possible to incorporate this in a performance indicator. After all, it is the road user who carries the responsibility to avoid accidents. But a road user needs assistance in making safe decisions. This need is answered by, for example, the Sustainable Safety vision (Wegman and Aarts, 2005) which is also the basis for the safe system approach (OECD/ITF, 2008).

5.3. Benchmarking background variables: structure and culture

The bottom layer of the pyramid, called structure and culture, has not yet been very well defined in the SUNflower approach. SUNflower added an extra layer to the model which was developed in New Zealand (LTSA, 2000). The reasons were twofold:

- It provides an essential basis for all the observations and indicators at a higher level of the pyramid. Progress in road safety could perhaps not be fully understood or even be misinterpreted by not knowing or ignoring these backgrounds.
- It is not easy to transfer findings of benchmarking and to learn from experiences and results in other countries without having a clear picture of the setting in which these results were achieved or the changes were measured.

In the Structure part of the bottom layer two dimensions are distinguished: physical structure and operational (functional) structure.

The physical structure of a country is described by numerous factors that can be defined as specific long-term conditions contributing to different road safety outcomes. They are typically not, or at least not only, amenable to interventions by conventional road safety policies. Moreover, they are typically modifiable by more general policies, in a long term only. The two groups of structural factors can be distinguished by their amenability to interventions in time (Eksler, 2009): (1) stationary factors – not changing in time (e.g. geographic and climate conditions) and (2) tractable (dynamic) factors – subject to evolutions or changes in time (e.g. demography, road topology, and urbanization).

The operational structure refers to the organisation of and arrangements between all potential actors involved in policy making in relation with road safety. Therefore, this is where the manner is discussed in which society uses institutions to try and solve social problems; road safety in this case. Somehow, there will need to be agreements between the various actors about their contribution to road safety improvement. If this cooperation is not well coordinated, loss of both quality and efficiency will be the result. The government's role in this is especially important, or rather the roles of the different layers of government.

Culture is defined by values and norms in their social sense. Values can be regarded as assumptions upon which implementations can be based. Sets of consistent values and measures together form the value system which is subjective and varies across people. Such values have an influence on behavioural attitudes of road users, which in turn will manifest themselves in different road safety outcomes. Social sanctioning is based on reference standards that regulate behaviour and act as informal social control. A striking example is society's attitude towards drink-driving, which differs significantly between countries.

Road safety culture is reflected in the way society deals with the consequences of the lack of road safety, to what extent these consequences are considered to be unavoidable, and the degree of social and political interest in eliminating or at least modifying these consequences. This road safety culture partly decides the political, governmental and social reactions to traffic risks and determines which 'price' one is willing to pay for risk reduction.

Countries' differences in 'structure and culture' not only have an effect on the size and nature of road safety problems, but they also influence the possibilities to reduce the problem effectively and efficiently. This is an important topic for future research.

6. Grouping countries: defining classes of countries?

Benchmarking is a process in which countries or jurisdictions evaluate various aspects of their performance in relation to other and so-called 'best-in-class' practices. Two approaches for identifying classes for benchmarking purposes are possible: to just put all countries in one class, or alternatively, to identify several classes and group comparable countries per class. Theoretically speaking, both approaches are valid. However, from a practical point of view, it is to be expected that countries can learn more easily from similar countries. This suggests that forming classes of more or less similar and comparable countries does make sense. And it could be argued that pupils in the same class are more motivated to make improvements if being the 'best-in-class' is considered to be within reach.

It is relevant to research the possibilities and consequences of grouping (European) countries in such a manner that comparisons can be made against more similar backgrounds. Three alternative ways to carry out such a grouping were tested.

The first grouping was carried out independently by a small number of traffic safety experts. The second grouping was made based on the level of safety and the safety development for which

data was available. The third grouping was made on a much larger number of social, economic and geographic characteristics. The first grouping differs from the other two, because it is not directly based on objective data. Pilot tests were carried out with all three ways of grouping and were compared with each other. These preliminary tests were carried out on existing data that was easily available. The objective was primarily to show how the existing techniques could be applied and how the results relate to each other.

The following grouping was suggested by the experts: Group 1: DK, FI, IS, IE, NL, NO, SE, UK; Group 2: AT, BE, FR, DE, IT, LU, CH; Group 3: CZ, HU, PL, SK, SL; Group 4: EL, PT, SP. In general, the grouping was consistent. Discussion was possible about Italy (in Group 2 or 4), Ireland (in Group 1 or 2), the Netherlands (in Group 1 or 2) and Slovenia (in Group 2 or 3).

The second grouping strategy used a Singular Value Decomposition (SVD) of the annual fatality risks in the years 1980–2003 (see for example Lay, 1994). The third strategy used a Multiple Correspondence Analysis (MCA) of a number of characteristics of the countries as observed in the most recent years (Gifi, 1990). The purpose of both the SVD and the MCA technique is to detect trends in characteristics that countries have in common: that is, to capture the most important and salient relations between the variables. In the SVD these are the developments in fatality risk, in the MCA they are the characteristics of the countries (the observational units) in just a few (e.g. two) dimensions. The main goal of the present analysis is to capture the most important similarities between countries. Both techniques produce values on two (or more) dimensions that can be visualized in a plot in which each country is represented by a dot; in this plot small distances between dots imply that the corresponding countries are quite similar to each other (in terms of the values of the variables used in the analysis), while large distances between dots represent countries that are very different from each other. Analysis of the distances between countries in this plot is therefore particularly suited for grouping countries, even though – as we will see – there is still a certain amount of subjectivity involved in the grouping of countries based on these plots.

Whether to use an SVD or an MCA for our purposes typically depends on the measurement level of the variables used in the analysis. When trends are analyzed on continuous numerical variables (such as fatality risks) SVD is the most appropriate technique. On the other hand, when we use discrete nominal variables of different character, such as religion or type of education, MCA is the appropriate technique.

The second (SVD) strategy used available data from 13 European countries. Fig. 3 shows the plotted factor scores on dimensions 1 and 2 for these countries. The first dimension is the best representation of the development for all the countries as it is a weighted average of all risk developments. The next dimension is the best representation of the residuals from the first dimension. This dimension represents the most important 'correction' of the first general trend.

Based on Fig. 3, the countries can be grouped as follows: the Nordic countries are located in the upper-left part of the plane, together with Great Britain and the Netherlands. To the right of this group and a bit lower Switzerland, Germany, France, Belgium and, a bit further away, Austria can be distinguished as a second group. Ellipses have been drawn only to indicate these two groups, suggesting a possible grouping of these countries. The choice for Switzerland in Group 2 is a bit arbitrary. Slovenia and Portugal have isolated positions because of their discontinuous risk developments in time.

The idea behind the third (MCA) strategy is that there are many factors other than traffic and safety characteristics that may influence the traffic system or restrict possibilities for safety measures. Examples are the geographical, social or cultural characteristics of a country or society. To compare a country's safety position with that in other countries, it is convenient to select those countries that are similar regarding these background variables. There is a wide variety of characteristics and they are not all directly measurable in the ordinary sense. Examples are the type of religion, literacy et cetera. An alternative to the Multiple Correspondence Analysis (MCA) is the classical Principal Components Analysis (PCA). This technique however is not suitable for rank order data or nominal variables.

Because the data are from various sources (a.o. ITF/OECD, Statistics Netherlands) of have a different nature and are not directly related to traffic safety, it is difficult to collect the necessary information for all European countries. To show how such a tech-

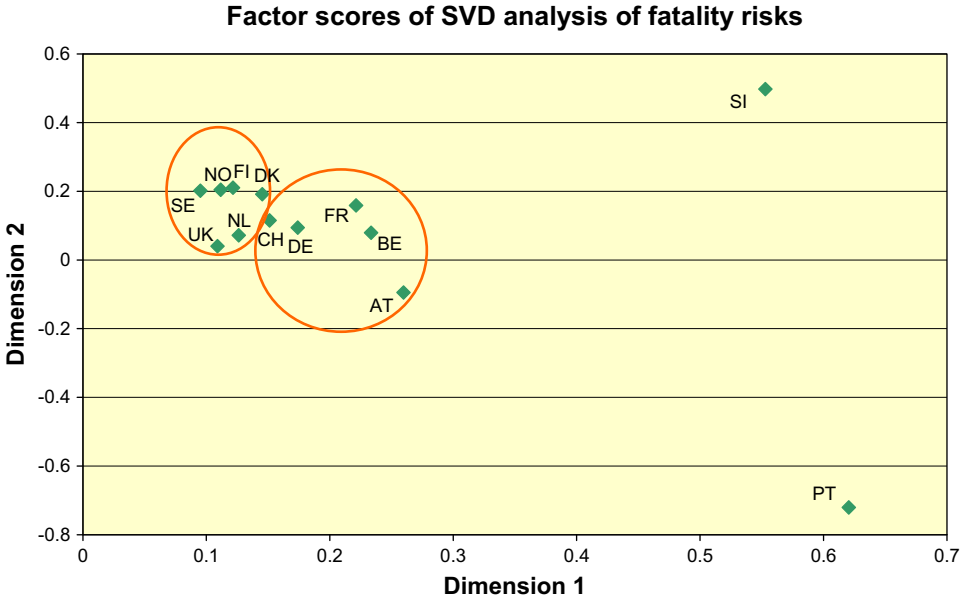

Fig. 3. Factor scores for 13 countries on the first two dimensions of a SVD analysis on fatality risks.

nique can be used for the purpose of grouping countries, ten readily available characteristics were used in the analysis (Wegman et al., 2008). A disadvantage of the choice that was made is the fact that many of the used characteristics are of the same type as or are strongly correlated with other characteristics. Examples are variables such as winter temperature, summer temperature, latitude and religion. The outcome of any MCA analysis will therefore be dominated by such clusters, because they are shown to be dominant underlying components. In a more balanced set of variables such artefacts disappear. Despite its drawbacks, the analysis is reported here to show its usefulness for grouping countries. The data used in the analysis was collected for 23 countries (see Fig. 4). The following ten variables were used: seatbelt use, mountainous areas, winter temperature, summer temperature, part of the population with at least secondary education, latitude, main religion, literacy, population density and, gross domestic product. Information on important issues such as the safety quality of the road network, road types and road use, vehicle fleet, modal split, road user education, safety measures taken and other safety performance indicators are not readily available for many countries.

The classification resulted in a matrix of 23 countries by 10 variables. The MCA of this data set was performed in two dimensions. The dimensions are abstract characteristics, representing the combinations of the original variables that best represent similarities between countries. The first dimension is once more the most important 'common trait' of the countries. The second dimension is the next important one, best describing the residuals that cannot be described by the first dimension. The two dimensions are independent from each other.

The results of this MCA are presented in Fig. 4.

The most important variables responsible for the clustering found in Fig. 4 are (non-linear transformations of) the categories of 'latitude', 'population density', and 'winter temperature'. The first dimension represents latitude primarily. The second dimension is more difficult to interpret, but population density is an important factor in this dimension. This result illustrates the importance of a careful selection of the variables used as the basis on which to form groups of countries.

The results of the three grouping strategies have many points of agreement especially regarding Groups 1 and 4 (Table 1). Therefore, this preliminary grouping is somewhat arbitrary. The results in the table are therefore presented with some reservation. Further effort is recommended to improve the quality of this grouping and to add more countries to this grouping.

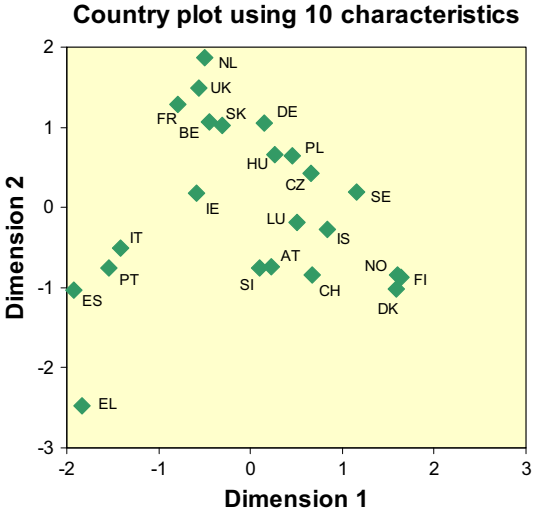

Fig. 4. Plot of the 23 countries in the two-dimensional plane, resulting from the MCA, using 10 characteristics of these countries.

Table 1
Indicative grouping of countries based on three different procedures to be used in further benchmarking of road safety performances.

Group 1	Group 2	Group 3	Group 4
Denmark	Austria	Czech Republic	Greece
Finland	Belgium	Hungary	Portugal
Great Britain	France	Poland	Spain
Iceland	Germany	Slovakia	
The Netherlands	Ireland	Slovenia	
Norway	Italy		
Sweden	Luxemburg		
	Switzerland		

To conclude: it can be argued that it is worthwhile to first group similar countries and to then compare the countries within a specific group ('class'). As three different procedures (expert opinion, road safety trends and general background data) resulted in 'many points of agreement' we conclude that rather than making comparisons between all European countries, it is better group countries first. Based on our results we conclude that such a grouping does not cause major problems.

7. Conclusions

Benchmarking the road safety performance of countries as a basis for learning and speeding up positive developments can be considered a promising step in improving road safety. Mainly because of the simplicity of the approach and its appeal to a wider audience, among whom politicians and policymakers, benchmarking is already applied in many fields. However, it has not really made its entrance into the world of road safety yet. A simple ranking, but even better, a well-accepted benchmarking could result in inviting experts to explain positions and to explain changes in positions to a wider audience. Without any doubt, such benchmark results will attract attention from the media and this can be used to make further steps. However, if we consider benchmarking mainly as a basis for learning from one another, we are not only interested in the final score/rate/ranking, but also in the backgrounds of those scores, in the components that contribute to the scores, and in the potential for improvements. Benchmark results need to be accepted by policy makers and scientists; this report has been written to obtain support in the field of road safety.

Benchmarking is a process in which countries or sub-national jurisdictions (regions and states), evaluate various aspects of their performance in relation to other practices, among which the so-called 'best-in-class'. The benchmark results enable countries or jurisdictions to learn from others as a basis for developing measures and programmes which are aimed at increasing their own performance.

To be able to carry out meaningful benchmarking, performance indicators need to be designed. The advantages of working with these indicators are imminent: simplification, quantification and communication. Wegman et al. (2008) distinguish three types of indicators covering all elements of the SUNflower-pyramid: road safety performance indicators, policy performance indicators and implementation performance indicators. The three types of indicators are embedded in a policy context, 'the structure and culture of a country', which are considered as background variables. The three indicators should be combined into a composite index. A comparison of the pros and cons leads to the conclusion that it is attractive to develop a composite index for road safety.

By developing a composite index for road safety (the SUNflower road safety performance index?) it has become apparent that we still lack the knowledge to include all relevant aspects and, in addition, we lack data. Nevertheless, it is considered feasible to study performance indicators based on information from all EU Member

States. It is recommended to develop valid and reliable indicators, both for policy performance and for implementation performance.

In order to evaluate effects of safety programmes it is important to compare not just the most recent situation but also safety developments over time between countries. To analyse these developments trend-analysis techniques are required and indicators proposed. Finally it is recommended to make comparisons and benchmarks not only between countries but also for regions within countries.

A few subjects for further research are identified throughout this paper. Detailing the different performance indicators is a key area for research. Furthermore, it is anticipated to collect compatible national data sets in a European research project and use these data to design a composite performance index.

Acknowledgements

We thank the colleagues we cooperated with in the several SUNflower projects, especially our colleagues participating in the SUNflowerNext project.

References

AAA, 2007. Improving Traffic Safety Culture in the United States. The Journey Forward. American Automobile Association AAA Foundation for Traffic Safety, Washington, DC.

Adriaanse, A., 1993. Environmental Policy Performance Indicators. A Study on the Development of Indicators for Environmental Policy in the Netherlands. Ministry of Housing, Physical Planning and the Environment. Sdu Publishers, The Hague.

Bliss, A., Breen, J.M., 2009. Implementing the Recommendations of the World Report on Road Traffic Injury Prevention. Country Guidelines for the Conduct of Road Safety Management Capacity Review and the Specification of Lead Agency Reforms, Investment Strategies and Safe System Projects. World Bank, Washington, DC.

Camp, R.C., 1989. Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance. Quality Press for the American Society for Quality Control, Milwaukee, Wisconsin.

Eksler, V., 2009. Road Mortality in the EU: A Regional Approach. PhD Thesis, Versailles St. Quentin University, Versailles.

Elvik, R., Veisten, K., 2005. Barriers to the Use of Efficiency Assessment Tools in Road Safety Policy. Workpackage 2 of the European Research Project ROSEBUD. TØI rapport 785/2005, Institute of Transport Economics, Oslo.

ETSC 2006. A methodological approach to national road safety policies. European Transport safety Council ETSC, Brussels.

Gifi, A., 1990. Nonlinear Multivariate Analysis. Wiley, New York.

Haddon Jr., W., 1968. The changing approach to the epidemiology, prevention, and amelioration of trauma: the transition to approaches etiologically rather than descriptively. American Journal of Public Health 1968 (58), 1431–1438.

Hermans, E, 2009. A Methodology for Developing a Composite Road Safety Performance Index for Cross-Country Comparison, PhD Thesis, Hasselt University, Hasselt, Belgium.

Koornstra, M., Lynam, D., Nilsson, G., Noordzij, P., Pettersson, H.-E., Wegman, F., Wouters, P., 2002. SUNflower. A Comparative Study of the Development of Road Safety in Sweden, the United Kingdom, and the Netherlands. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

Lay, D.C., 1994. Linear Algebra and Its Applications. Addison-Wesley Publishing Company, New York.

LTSA, 2000. Road Safety Strategy 2010. A Consultation Document. National Road Safety Committee, Land Transport Safety Authority LTSA, Wellington, New Zealand.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2005. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Statistics Working Papers, 2005/3, Organisation for Economic Co-operation and Development OECD, Paris.

OECD, 2002. Safety on roads. What's the vision? Organisation for Economic Co-operation and Development OECD, Paris.

OECD/ITF, 2008. Towards Zero – Ambitious Road Safety Targets and the Safe System Approach. Joint Transport Research Centre of the OECD and the International Transport Forum, Paris.

OECD/JRC, 2008. Handbook on Constructing Composite Indicators. Methodology and User Guide. Organisation for Economic Co-operation and Development OECD and the Joint Research Centre of the European Commission. OECD, Paris.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A.A., Jarawan, E., Mathers, C., 2004. World Report on Road Traffic Injury Prevention. World Health Organization (WHO), Geneva.

Saisana, M., Tarantola, S., 2002. State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development. EUR 20408 EN. Joint Research Centre of the European Commission, Italy.

Wegman, F., 2004. Implementing, Monitoring, Evaluating and Updating a Road Safety Programme. D-2003–12. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

Wegman, F., Aarts, L. (Eds.), 2005. Advancing Sustainable Safety. National Road Safety Outlook for 2020. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

Wegman, F., Eksler, V., Hayes, S., Lynam, D., Morsink, P., Oppe, S., 2005. SUNflower+6. A Comparative Study of the Development of Road Safety in the SUNflower+6 Countries: Final Report. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

Wegman, F., Commandeur, J., Doveh, E., Eksler, V., Gitelman, V., Hakkert, S., Lynam, D., Oppe, S., 2008. SUNflowerNext: Towards a Composite Performance Index. SWOV, Leidschendam, The Netherlands.