

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Combining road safety information in a performance index

Elke Hermans^a, Filip Van den Bossche^{b,1}, Geert Wets^{a,*}^a Hasselt University, Transportation Research Institute, Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium^b European University College, Research Group Economics and Policy, Stormstraat 2, 1000 Brussels, Belgium

ARTICLE INFO

Article history:

Received 29 August 2007

Received in revised form 19 November 2007

Accepted 14 February 2008

Keywords:

Road safety
Performance index
Weighting methods
Indicators

ABSTRACT

In this paper we focus on an essential step in the construction process of a composite road safety performance indicator: the assignment of weights to the individual indicators. In the composite indicator literature, this subject has been discussed for a long time, and no agreement has been reached so far. The aim of this research is to provide insights in the most important weighting methods: factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting. We will give the essential theoretical considerations, apply the methods on road safety data from various countries and discuss their advantages and disadvantages. This will facilitate the selection of a justifiable method. It is shown that the position of a country in the ranking is influenced by the method used. The weighting methods agree more for countries with a relatively bad road safety performance. Of the five techniques, the weights based on data envelopment analysis resulted in the highest correlation with the road safety ranking of 21 European countries based on the number of traffic fatalities per million inhabitants. This method is valuable for the development of a road safety index.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The concept of indicators has gained popularity in recent years. In general, an indicator can be defined as a quantitative or qualitative measure that is deduced from a series of observed facts to reveal the relative positions of objects in a certain area (Nardo et al., 2005). One of the useful characteristics is that an indicator can represent large amounts of information in a simple manner. Indicators can be used for several objectives, like monitoring performance, identifying trends, predicting problems, assessing policy impact, prioritizing measures, benchmarking, etc. (Litman, 2007; Sharpe, 2004).

Moreover, there is a rapid development of composite indicators – or indexes which are a combination of individual indicators – in several domains. Examples are the Human Development Index, the Technology Achievement Index and the Environmental Sustainability Index (Saisana and Tarantola, 2002). In order to combine information from several underlying dimensions into one index, an essential step is to assign a correct weight to each indicator. That way, policymakers have a useful tool for prioritising their actions. In the literature, some attention has already been paid to the issue of weighting (Nardo et al., 2005; Sharpe, 2004). In general, opin-

ions are different regarding this aspect, resulting in two groups. The one in favour of aggregating information believe that a summary statistic can capture reality, is meaningful and attracts attention, while opponents claim that no aggregation should be performed because of what they call the arbitrary nature of the weighting process (Sharpe, 2004).

Road safety is one of the policy areas where the use of indicators is rapidly gaining ground (e.g. SafetyNet, 2005). The multidisciplinary character of road safety implies that policymakers should take various influential factors into account. Also, being an international issue, the level of road safety is often compared over countries. To reduce the dimensions of the problem, the creation of a composite indicator can be helpful. One of the main issues in creating a road safety performance index is the weighting of the indicators. This exercise implies that a value judgement is needed for each of the possible measures taken to influence road safety. A higher weight for a certain indicator stresses its relatively higher importance in the global measurement of road safety. Given the limited resources that can be used to improve road safety, a well-considered construction of indicator weights is crucial to steer future investments.

In our opinion, the advantages of combining road safety information can be decisive, given that a sound weighting framework is created. Therefore, we focus in this paper on the development of a composite road safety indicator for European countries. That is, our aim is to assign to a country an index score that captures all relevant information concerning the road safety situation in order to benchmark the performance of that country. Despite the recent

* Corresponding author. Tel.: +32 2 608 14 17; fax: +32 11 26 91 99.

E-mail addresses: elke.hermans@uhasselt.be (E. Hermans),filip.vandenbossche@ehsal.be (F. Van den Bossche), geert.wets@uhasselt.be (G. Wets).¹ Tel.: +32 2 210 14 17.

interest of the road safety domain in indicators, the development of one road safety index is a new and challenging matter.

A number of steps are involved in the creation of a composite indicator (Nardo et al., 2005). The theoretical framework has to be developed, appropriate indicators selected and data found. Next, these values are weighted, aggregated and clearly presented. We start by translating the multidimensional road safety problem into a number of relevant factors related to accidents and casualties. Research regarding the identification of risk factors has already been carried out (Al Haji, 2005; WHO, 2004). Although the possibilities of risk factors are numerous, a safety performance indicator or SPI – as defined in ETSC (2001) and used on the European level in the SafetyNet project (SWOV, 2007), as well as the road safety performance index programme (ETSC, 2007) – is a measurement that is causally related to unsafety and used in addition to counting accidents or casualties, in order to indicate the safety performance or to understand the process that leads to accidents. Contrary to accident statistics and the related disadvantages – random fluctuation, under-registration, lack of uniformity in definitions, etc. – safety performance indicators are situated on the level of the underlying processes that cause accidents and casualties. SPIs give a more complete image, are able to detect problems before they result in accidents, are available earlier, etc. (ETSC, 2001). A large number of potential SPIs exist. The importance of the different indicator candidates depends on the strength of the relationship with unsafety, the amount of contribution to the accidents and the degree to which the factor can be influenced by policy measures (ETSC, 2001). The latter characteristic eliminates important road safety risk factors such as the weather (Eisenberg, 2004). In general, road safety performance indicators are related to the road user (e.g. speeding), the vehicle (e.g. defects) and the road (e.g. bad maintenance) (WHO, 2004).

This paper is organised as follows. In Section 2, the selected indicators that will be used to illustrate the different weighting methods are briefly discussed. The next section gives an overview of the characteristics of the five most often used weighting techniques. For factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting, a theoretical description of the technique is given and its strengths and weaknesses discussed. Subsequently, the five methods are applied to the road safety data set. The evaluation is the focus of Section 4 which compares the methods and deals with the robustness of the final results. The paper ends with conclusions and topics for further research.

2. Data

In the European SafetyNet project (2005) on safety performance indicators, the following seven domains are agreed on to be the most important road safety risk areas: alcohol and drugs; speed; protective systems; visibility; vehicle; infrastructure; and trauma care. For each risk domain, at least one appropriate indicator needs to be defined. Indicators for every risk domain are selected based on policy relevance, data availability, clarity and reliability (Litman, 2007; Nardo et al., 2005). To obtain an overall road safety index, two separate aggregation steps are to be performed. First, the individual indicators per risk domain are aggregated into one indicator per domain. Several possible indicators for each of the seven risk domains are presented in Fig. 1. Next, the domain indicators are aggregated in one road safety index. In this paper, it is assumed that each road safety risk domain can be fully represented by one carefully selected indicator. That is, the analysis is limited to the second aggregation step, in order to avoid unnecessary data complexities and to focus on the problem of weighting various risk domains in

one safety performance index. However, the presented framework can easily be extended to allow more indicators per risk domain.

For 21 European countries values for the 7 selected indicators are available from several international sources, amongst others the IRTAD international road traffic accident database, Eurostat, the World Health Organization and the SARTRE project on social attitudes to road traffic risk in Europe. As the indicators are expressed in different measurement units, we will use the standardised indicator values, as presented in Table 1. All indicator values can be interpreted in the same way: a higher value implies a higher level of road safety.

3. Weighting methods

In this section, the theory behind five common weighting techniques is discussed. We describe the working method and focus on the advantages and disadvantages of, respectively, factor analysis (Section 3.1), analytic hierarchy process (Section 3.2), budget allocation (Section 3.3), data envelopment analysis (Section 3.4) and equal weighting (Section 3.5). The application to the road safety data is the subject of Section 4.1.

3.1. Factor analysis

The first weighting procedure is based on factor analysis (FA). A factor analysis is often used to reduce the dimensions of a problem. In our road safety example there are 7 (l) dimensions explaining 100% of the variance of the problem. However, it would be interesting to reduce the problem to a smaller number of dimensions ($p < l$), called factors, which explain a large part of the total variance. Each factor consisting of a number of indicators can be given an interpretation. Several guidelines are available for assessing the optimal number of factors to which the problem can be reduced (Sharma, 1996). If the optimal number of factors is three or less, a graphical presentation can provide some useful insights. Having decided to consider p factors, rotation is a next step in order to enhance the interpretability. This results in each indicator (i) having a large factor score (a_{ij}) on one of the factors (j) only. The indicator weights can be deduced from these rotated factor loadings by means of relatively limited computation.

The procedure of deducing indicator weights from the rotated factor scores a_{ij} ($i = 1, \dots, l; j = 1, \dots, p$) is a simplified but equivalent approach (compared to the calculations suggested in Nardo et al., 2005) and consists of the following steps. Define $u_{ij} = a_{ij}^2 / \sum_{m=1}^l \sum_{n=1}^p a_{mn}^2$. The preliminary weight of indicator i , $u_i = \max_j(u_{ij})$. However, $U = \sum_i u_i < 1$ due to the reduction of the problem, leaving a small part of the variance unexplained. The final weight for each indicator i is equal to $w_i = u_i / U$ and, by construction, $W = \sum_i w_i = 1$.

The use of factor analysis in the composite indicators field (e.g. the e-Business Readiness Index (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 (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). This method is most valuable in case several indicators are considered to measure each risk domain.

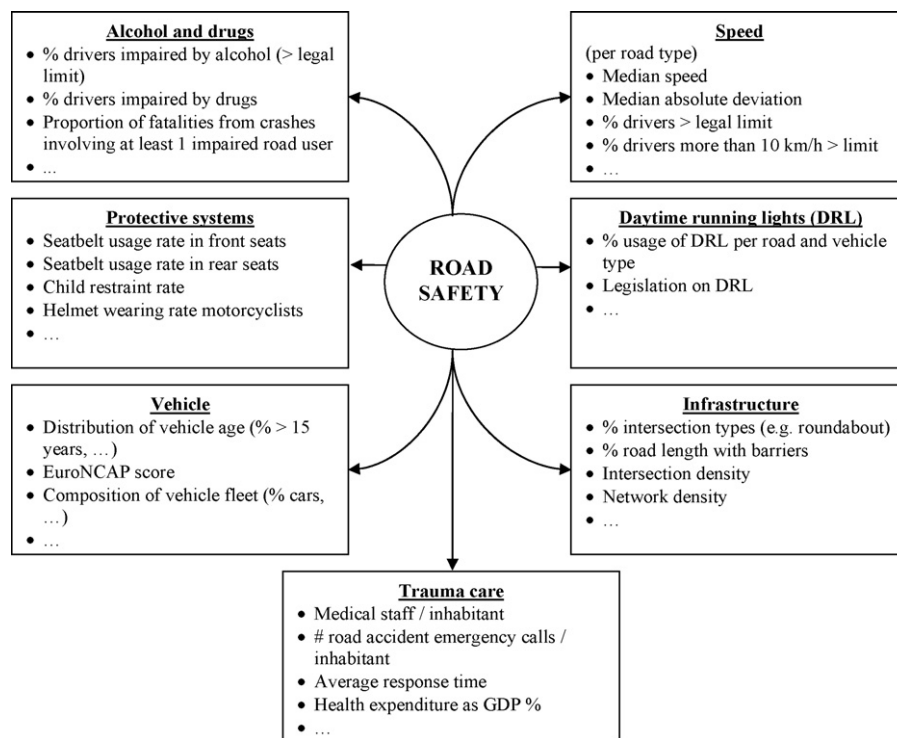


Fig. 1. Schematic overview of road safety indicators.

3.2. Analytic hierarchy process

Analytic hierarchy process or AHP is a method developed by Saaty in the early 1970s in the field of decision theory (Haas and Meixner, 2006). As stated by its name, a complex problem is translated into a hierarchy consisting of an overall goal (enhancing road safety), several (sub)criteria contributing to this goal and a number of alternatives of which the best has to be selected. Both quantitative and qualitative criteria can be handled (Haas and Meixner, 2006). In our context, experts judge the relative contribution of each indicator to road safety compared to another indicator. They answer the questions 'which one of the two is more contributing to the overall goal?' and 'how large is the intensity of the difference?'. Values are given on a scale of 1–9. 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 l indicators to judge, only $l(l-1)/2$ pairs have to be considered. The expert information can be presented in a reciprocal squared matrix with l rows and l columns consisting of values within the interval $[1/9, 1/8, \dots, 1, 2, \dots, 9]$.

From the matrix based on 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 of a matrix reflects the soundness of judgement, whether the interdependencies of the criteria are understood, etc. (Talbert et al., 1994). It is advisable to keep the number of criteria small and to define independent or at least sufficiently different criteria in each level of the hierarchy. Saaty (1980) defines the consistency index as $(\lambda_{\max} - l)/(l - 1)$. In addition, the consistency ratio is the ratio of the consistency index to the average random index for a matrix of the same size (which figure is given in Saaty, 1980). A consistency ratio equal to zero corresponds to a perfectly consistent matrix 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 sat-

isfactory consistency. Such small ratios do not drastically affect the weights (Nardo et al., 2005; Saaty, 1980).

Besides inconsistency, subjectivity is a characteristic of the AHP method making the team of experts crucial. Therefore, the selection should be well considered. 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 can be used for very complex decisions involving numerous levels of criteria and sub-criteria. It is a valuable technique for assessing indicator weights which has already been used in the Indoor Environment Index (Chiang and Lai, 2002) and the Index of Environmental Friendliness (Puolamaa et al., 1996). Despite the subjective characteristic and a possibly large inconsistency, the information from well-selected experts is valuable for deducing indicator weights. In case of several experts, numerous possibilities exist to come to one final set of weights. In the end, we want to obtain one weight for each indicator. This set of weights can be similar to the weights of one randomly selected expert or incorporate the opinion of 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.

3.3. Budget allocation

Budget allocation (BA) is another well-known 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 (s)he wants to stress its importance. The weights can be obtained from a simple ratio.

In general, the BA method has four phases (Nardo et al., 2005). 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 indi-

Table 1
Data set: standardised indicator values

Country	Alcohol and drugs % road users < BAC ^a limit	Speed % drivers < speed limit	Protective systems Seatbelt wearing rate in front	Visibility DRL ^b law	Vehicle % cars < 6 years	Infrastructure Network density	Trauma care Health expenditure as GDP ^c %
Austria	0.27	0.54	-0.13	1.29	0.10	0.00	-0.40
Belgium	-0.40	-0.17	-0.93	-0.06	0.67	3.45	0.54
Cyprus	-3.75	-1.12	0.09	-1.42	-1.68	-0.32	-1.27
Czech Republic	0.40	1.49	-1.29	-0.06	-1.45	0.03	-0.46
Denmark	0.75	-0.17	0.38	1.29	0.50	0.08	0.34
Estonia	0.48	-1.60	-0.28	1.29	-1.68	-0.36	-2.14
Finland	0.75	0.78	0.74	1.29	-0.50	-1.41	-0.66
France	-0.25	0.78	1.32	-0.06	0.30	0.23	0.94
Germany	0.32	-0.88	1.10	-0.06	0.37	0.22	1.75
Greece	-0.84	-1.36	-2.81	-1.42	-0.62	-0.73	0.81
Hungary	0.55	-0.88	-1.43	-0.06	-0.77	0.15	-0.33
Ireland	0.29	2.21	0.45	-1.42	2.01	-0.24	-0.66
Italy	-0.71	-0.41	-0.56	-0.06	0.11	0.02	0.14
Netherlands	0.42	-0.17	0.52	-1.42	0.47	1.51	0.54
Poland	0.75	0.54	-0.56	-0.06	-1.17	-0.24	-1.47
Portugal	-0.06	-0.41	0.66	-0.06	-0.53	-0.74	0.68
Slovenia	0.25	0.78	0.16	1.29	1.58	0.31	-0.06
Spain	-0.69	0.07	0.52	-0.06	0.20	-0.91	-0.46
Sweden	0.82	-0.17	0.95	1.29	0.53	-1.15	0.61
Switzerland	-0.06	-1.12	0.09	-0.06	0.55	0.15	1.95
United Kingdom	0.69	1.26	1.03	-1.42	1.02	-0.06	-0.40

^a Blood alcohol content.

^b Daytime running lights.

^c Gross domestic product.

icator 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 (e.g. the e-Business Readiness Index (Pennoni et al., 2005) and the Internal Market Index (Tarantola et al., 2004)) with some limitations. First, 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. For our road safety case experts from several countries were chosen to get a general view of the contribution of each indicator (domain) to road safety. Secondly, the method may not measure the importance of a specific indicator but the need for political intervention in that dimension (Nardo et al., 2005). Finally, the maximum number of indicators to distribute the budget over is limited to 10 enabling the expert to keep an overview (Saisana and Tarantola, 2002).

3.4. 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 (DMU's). 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, a set of weights is determined resulting in the best possible score for that country. This implies that dimensions on which the country performs relatively well get a higher weight.

Translating the original DEA context to the composite indicators 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). Consider y_{ij} as the raw value of indicator i ($i = 1, \dots, l$) for country j ($j = 1, \dots, n$). If $y_{ij} > y_{ik}$ then country j performs better than country k for that indicator. The constrained maximization problem can then be written as follows:

$$\begin{aligned}
 Cl_j = & \max_{w_i} \sum_{i=1}^l y_{ij} w_i \\
 \text{Subject to} & \\
 & \sum_{i=1}^l y_{ik} w_i \leq 1 \quad \forall k = 1, \dots, n \\
 & w_i \geq 0 \quad \forall i = 1, \dots, l
 \end{aligned} \tag{1}$$

This linear programming problem results in a set of indicator weights w_i for each country that maximizes the resulting composite indicator value Cl_j (i.e. the weighted sum of the indicators). As stated by the second constraint, all weights are restricted to be non-negative. The first constraint guarantees an intuitive interpretation of the composite indicator and implies that no country can be assigned an index value larger than 1 under these weights. For each country j we will obtain a composite index between zero and one, with higher values indicating a better relative performance.

For countries with an index score smaller than one, the country-specific weights can identify the sources and the amount of inefficiency in each indicator (Cooper et al., 2000). If the inefficiency of a country is reduced most in case of a better alcohol performance (that country has a low weight attached to alcohol), this information can be translated in specific alcohol action plans. Additionally, the indicator values of the efficient countries can be used to set realistic and useful targets for the inefficient country and the achievement towards these targets can be studied.

Compared to the previously discussed weighting methods, DEA is different. DEA is a method that can handle raw values making the normalisation of indicators redundant. However, this implies that the weights do not sum up to one, which makes the compari-

son of indicator weights with other weighting 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. As it is only correct to rank countries based on the same set of weights (see Adler et al., 2002) we will define one model for the whole set of countries in which the sum of all 21 composite indicator values has to be maximized.

Another adaptation to model (1) is the inclusion of extra restrictions on the weights. Total flexibility for the weights has been criticized on several grounds (Pedraja-Chaparro et al., 1997). The above model results in the maximum index score for 15 countries. To obtain a realistic ranking a multitude of ways to capture value judgements or prior information in DEA is proposed in the literature (e.g. Allen and Thanassoulis, 2004; Allen et al., 1997). Here, the contribution of each indicator pie share (i.e. the product of the indicator value and the indicator weight) to the composite indicator value is constrained to lie in the range defined by the minimum and maximum weight obtained from the budget allocation results over all experts (Cherchye et al., 2006). To get more robust restrictions we incorporate the opinion of four experts instead of two and take the average of the two highest weights (U_m) and the average of the two lowest weights (L_m) for each indicator. Algebraically, the DEA model that we will use is as follows:

$$\sum_j C_l_j = \max_{w_i} \sum_{i=1}^l \sum_{j=1}^n y_{ij} w_i$$

Subject to

$$\sum_{i=1}^l y_{ik} w_i \leq 1 \quad \forall k = 1, \dots, n \quad (2)$$

$$L_m \leq \frac{y_{mj} w_m}{\sum_{i=1}^l y_{ij} w_i} \leq U_m \quad \forall m = 1, \dots, l; \quad \forall j = 1, \dots, n$$

$$w_i \geq 0 \quad \forall i = 1, \dots, l$$

To conclude, DEA compares the performance of a country to the performance of the other countries in the data set. Since the

model chooses the optimal weights under the imposed restrictions, no other weighting set yields a higher composite indicator value. The results are influenced by the countries in the data set, hence this approach is only about relative efficiency (Anderson, 2006). Although the DEA domain is very extensive and numerous models exist, making the selection of an appropriate model a rather difficult task, this technique can be translated to be useful for composite indicators. This weighting method has already been used for a number of indexes (Cherchye et al., 2004, 2006; Cherchye and Kuosmanen, 2002) and to a limited extent in the road safety context (Odeck, 2005). Its strongest point is that the weights are endogenously determined and derived directly from the data (Cooper et al., 2000). However, a number of disadvantages are linked to the standard model leading us to adapt some aspects. As we developed one model for all countries in the data set, not only the computational effort is reduced but also one set of indicator weights is obtained rendering the ranking of countries meaningful. Furthermore, bounds defined by more than one expert can easily be incorporated leading to more acceptable weights. The presented DEA model is most valuable when individual expert opinions are available and there is no agreement on the correct set of weights.

3.5. Equal weighting

As is clear by the name of this method, the same weight is assigned to each indicator. Since the sum of all weights equals one, each indicator gets the weight $1/l$ (with l the number of indicators in the analysis). It is also possible to use equal weighting for the main categories and for all indicators in those categories. In that case the weights depend on the number of indicators in each category.

Although from a scientific point of view equal weighting is a too simple technique, a large majority of composite indicators is constructed by means of this default weighting method, e.g. the European Innovation Scoreboard (CEC, 2004) and the Environmental Sustainability Index (Yale, 2005). The most important drawback is that no insights are gained in 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

Table 2
Summary information on the five weighting methods

	Main advantages	Main disadvantages	Main requirements
Factor analysis	Indicators are grouped An interpretation can be given to each factor	Weights based on correlations may differ from reality	Some correlation between indicators Justification of the optimal number of factors Clear rotation results
Analytic hierarchy process	Detailed expert information Incorporate quantitative and qualitative criteria Numerous levels of criteria	Inconsistency Subjectivity	Carefully selected group of experts (with time) Small number of criteria Sufficiently different criteria
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 10
Data envelopment analysis	Optimal weights derived from data and restrictions 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 (like for other weighting methods)	Value judgements to obtain realistic weights Several countries in data set
Equal weighting	Simple	No insights 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

Table 3
Four sets of road safety indicator weights and the average indicator contribution (*) of the data envelopment analysis

	FA	AHP	BA	DEA*	EW
Alcohol	0.139	0.229	0.234	0.259	0.143
Speed	0.190	0.262	0.305	0.250	0.143
Protective systems	0.095	0.129	0.132	0.211	0.143
Visibility	0.163	0.028	0.043	0.000	0.143
Vehicle	0.100	0.102	0.071	0.035	0.143
Infrastructure	0.139	0.179	0.105	0.080	0.143
Trauma care	0.174	0.071	0.110	0.028	0.143

resulting weights are similar to the real, unknown weights. When two or more indicators are measuring the same, there is a risk of double weighting (DSTI, 2003). Finally, in case of fixed weighting, country rankings may change by using another normalisation method (Cherchye et al., 2006). 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.

4. Results

The previous section was devoted to the description of five common and useful weighting methods for combining information in one index, namely factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting. Table 2 summarizes the main advantages, disadvantages and requirements of each method. In this section we apply the methods to the road safety data presented in Table 1 and evaluate the methods by comparing their results in terms of weights and rankings.

4.1. Application of the weighting methods

Factor analysis was applied to the road safety data set in order to obtain a first set of indicator weights. The optimal number of factors for the data presented in Table 1 proves to be 4. These four factors explain more than 83% of the total variance. Factor 1 consists of speed. Vehicle and trauma care are linked to the second factor. Alcohol and visibility have their highest loading on factor 3 while the protective systems indicator and the infrastructure indicator belong to factor 4. The (7 × 4) rotated factor loadings from the output were used to find the indicator weights, which are presented in the second column of Table 3.

Secondly, weights were obtained using the analytic hierarchy process. The first step was to gather a group of experts for judging the relative contribution of the indicators. For our road safety case we obtained judgements from 12 experts originating from 9 European countries, all having a knowledge of the road safety field. They were asked to give their own opinion independently from others. Using the Expert Choice software (Expert Choice Inc., 2006) the 7 indicator weights were determined out of each of the 12 matrices and the consistency ratios computed by dividing each consistency index by 1.32 (Saaty, 1980). There are several options to obtain one set of weights. For the purpose of this study, we incorporate the available knowledge of the experts with an acceptable degree of consistency (smaller than 0.10) and take the average over eight sets of weights. The indicator weights based on AHP are presented in column 3 of Table 3.

Thirdly, we obtained valid budget allocation results from 11 road safety experts out of 9 European countries. Each expert had to distribute a budget of 70,000 € over the 7 road safety domains. In accordance with the analytic hierarchy process, we consider the mean indicator weights shown in the fourth column of Table 3.

Data envelopment analysis is the fourth method under study. Several DEA software packages exist (see e.g. Barr, 2004). In this

paper we opt for the DEA Excel Solver program with its extensive collection of models and appealing user interface. We applied the second data envelopment analysis model specified in Section 3.4 to our road safety data set (before normalisation). For each country, the contribution of each indicator to road safety is given in Table 4 with the last column containing the composite indicator values, which are the sum of the 7 pie shares for each country.

The sum of the composite indicator scores is 18.13. Only one country – Belgium – obtained the maximum score due to its very high infrastructure value. Although all seven indicator domains are considered very relevant for road safety, the visibility domain (and the daytime running lights indicator) receives zero weight. Since two experts distributed zero budget to this indicator in the budget allocation exercise, this is justified in a way. Based on the composite indicator scores all countries can be ranked.

Fifthly, the equal weighting method assigned to each of the 7 road safety indicators a weight of 1/7 or 0.143. These values are shown in column 6 of Table 3.

4.2. Discussion

Of the five considered methods data envelopment analysis is the most different technique. For this method, instead of seven indicator weights summing up to one, the average indicator contribution to road safety over all countries is presented in Table 3. 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 hand. Equivalent to equal weighting, the spread in the factor analysis weights is rather limited. Speed, visibility and trauma care obtain a relatively higher weight while protective systems and vehicle are assigned less weight. As could be intuitively expected, the resulting weights from the participatory methods AHP and BA using the same group of experts are alike. However, it can be noticed that the way of formulating the question may lead to different results as is clearly illustrated by infrastructure. In general, alcohol and speed are indicated as the most important risk determinants of road safety while visibility, vehicle and trauma care are less influencing according to the experts.

Each set of weights is used to calculate the road safety index score for the 21 countries. We compute the sum of the product of each (normalised) indicator value and its assigned weight to obtain these scores and to rank the countries in such a manner that a higher road safety index score relates to a higher rank. In other words, rank 1 represents the most road safe country taking into account the indicator values and the assigned weights. The rankings resulting from the five weighting methods even as the road safety (RS) ranking based on the number of traffic fatalities per million inhabitants are shown in Table 5. This normalised road safety ranking is frequently used to compare countries differing in size on their level of road safety. Note that this ranking is not based on a composite index but on one outcome indicator (which provides no explanation for the process leading to an accident). In case of an equal number of

Table 4
Road safety indicator contributions and composite indicator values based on data envelopment analysis and (budget allocation) expert information

	Alcohol	Speed	Protective systems	Visibility	Vehicle	Infrastructure	Trauma care	CI
Austria	0.262	0.257	0.206	0.000	0.036	0.081	0.026	0.868
Belgium	0.254	0.248	0.176	0.000	0.042	0.249	0.031	1.000
Cyprus	0.211	0.237	0.214	0.000	0.017	0.065	0.022	0.765
Czech Republic	0.264	0.269	0.163	0.000	0.020	0.082	0.026	0.823
Denmark	0.269	0.248	0.224	0.000	0.040	0.085	0.030	0.896
Estonia	0.265	0.231	0.200	0.000	0.017	0.063	0.017	0.794
Finland	0.269	0.260	0.238	0.000	0.030	0.012	0.025	0.832
France	0.256	0.260	0.259	0.000	0.038	0.092	0.033	0.938
Germany	0.263	0.240	0.251	0.000	0.039	0.091	0.037	0.921
Greece	0.248	0.234	0.107	0.000	0.028	0.045	0.033	0.695
Hungary	0.266	0.240	0.158	0.000	0.027	0.088	0.027	0.804
Ireland	0.263	0.277	0.227	0.000	0.056	0.069	0.025	0.916
Italy	0.250	0.245	0.190	0.000	0.036	0.082	0.029	0.831
Netherlands	0.264	0.248	0.230	0.000	0.040	0.154	0.031	0.967
Poland	0.269	0.257	0.190	0.000	0.023	0.069	0.021	0.828
Portugal	0.258	0.245	0.235	0.000	0.029	0.045	0.032	0.844
Slovenia	0.262	0.260	0.216	0.000	0.051	0.096	0.028	0.913
Spain	0.250	0.251	0.230	0.000	0.037	0.036	0.026	0.830
Sweden	0.269	0.248	0.246	0.000	0.040	0.024	0.032	0.860
Switzerland	0.258	0.237	0.214	0.000	0.040	0.088	0.038	0.875
United Kingdom	0.268	0.266	0.248	0.000	0.045	0.078	0.026	0.931

traffic fatalities per million inhabitants the average rank is taken for both countries.

Depending on the situation and the available information, a researcher can opt for one of the weighting methods described. However, we also want to express which method is to be preferred in this case. The rankings resulting from the five weighting methods are compared based on the knowledge of the road safety domain and the specific characteristics of the weighting methods. In general, we use indicators to describe the road safety performance of a country and the road safety index – a weighted sum of indicators – should be a good approximation of the number of road fatalities per million inhabitants. Therefore, a correlation analysis on each weighting ranking and the road safety ranking is performed. In this case, the ranking of the countries based on the data envelopment analysis has the best fit with the ranking using the number of traffic fatalities per million inhabitants. In other words, given the indicator values and the imposed restrictions the optimisation algorithm of DEA results in the best resemblance with the road safety ranking.

The behavioural indicators – alcohol, speed and protective systems – have a large share in the overall road safety index.

A closer look at Table 5 shows that Slovenia and France have the highest average rank over the five weighting methods. As can be deduced from their indicator values in Table 1, these two countries have a top-5 rank number on four indicators and no bottom-5 rank number. Furthermore, the five weighting methods agree most on the ranking of Estonia, Hungary, Cyprus and Greece and also on Spain and Austria. These countries have a low road safety rank regardless of the weighting method used. Belgium has a bad rank under the budget allocation method. The data in Table 1 show that Belgium scores relatively well on the dimensions vehicle and infrastructure. BA assigns the lowest weight to these dimensions compared to FA, AHP and EW (remember that for DEA the average contributions are given in Table 3 instead of the weights and Table 4 shows that the contribution of infrastructure for Belgium is 0.249). Its high infrastructure value results in the best rank under the DEA method. The same applies to the Netherlands. Denmark

Table 5
Rankings based on the five weighting methods and the number of traffic fatalities per million inhabitants

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

and Sweden obtain a good rank under the FA and EW ranking. This can be partly explained by their good visibility values and relatively bad performance regarding speed. Ireland is favoured by a low weight assigned to trauma care and visibility (on which it scores badly) and gets the highest rank using the AHP method. Finally, the favourable alcohol and speed values of Finland are enhanced by the high weights assigned to them by the budget allocation method.

5. Conclusions and further research

In this paper, five common methods for assigning weights to indicators are discussed, namely factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting. The theoretical considerations are described as well as the advantages, the disadvantages and the results of the application of each method to the road safety data set consisting of 7 road safety indicator values for 21 European countries.

From each method, a different set of indicator weights was obtained resulting in a relatively higher or lower contribution of an indicator to the overall aggregated index (e.g. AHP and BA assigned the largest impact to alcohol and speed). Consequently, the countries' ranking changed to some extent (the average difference in rank over the 21 countries based on the 5 weighting techniques equals 2.2). Countries with a relatively bad road safety performance (such as Greece, Cyprus, Hungary and Estonia) were less affected by the choice of the weighting method while a shift in rank at the top occurs more often (United Kingdom, the Netherlands, Belgium, Germany and Ireland).

Of the five methods, the data envelopment analysis method resulted in a ranking which best approaches the road safety ranking based on the number of traffic fatalities per million inhabitants. This method is obviously valuable in the road safety context as the original DEA idea can be adapted resulting in a useful optimisation problem. That way, acceptable and realistic weights are obtained for each indicator. As the weights are derived from the data, they are only influenced by the indicator values and the imposed restrictions.

The behavioural indicators – alcohol, speed and protective systems – have a large share in the overall road safety index. An improvement in these aspects will increase the road safety index score of a country the most. In the future, the data envelopment analysis method will be extended. In particular, a country-specific analysis will provide valuable road safety enhancing recommendations. For each country, risk domains that need urgent action can be identified by means of relatively low weights; for each country with an index score lower than one a set of relevant benchmark countries will be determined; a priority ranking can be given to the risk domains, etc.

Besides weighting, attention needs to be paid to the other aspects in the development process of a road safety performance index such as indicator selection, imputation of missing values and aggregation. In the end, we want to obtain an index on which policy decisions can be based, to describe and explain developments in performance over time, etc. To this end, several methods need to be considered during the different phases and a robustness analysis is required.

References

- Adler, N., Friedman, L., Sinuany-Stern, Z., 2002. Review of ranking methods in the data envelopment analysis context. *Eur. J. Oper. Res.* 140, 249–265.
- Al Haji, G., 2005. Towards A Road Safety Development Index. Linköpings universitet, Norrköping.
- Allen, R., Athanassopoulos, A., Dyson, R.G., Thanassoulis, E., 1997. Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Ann. Oper. Res.* 73, 13–34.
- Allen, R., Thanassoulis, E., 2004. Improving envelopment in data envelopment analysis. *Eur. J. Oper. Res.* 154 (2), 363–379.
- Anderson, T., 2006. A data envelopment analysis home page. Retrieved August 25, 2006, from <http://www.etm.pdx.edu/dea/homedeia.html>.
- Barr, R.S., 2004. DEA Software Tools and Technology: A State-of-the-Art Survey. Southern Methodist University.
- Commission of the European Communities (CEC), 2004. European Innovation Scoreboard 2004: Comparative Analysis of Innovation Performance.
- Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.* 2, 429–444.
- Cherchye, L., Kuosmanen, T., 2002. Benchmarking Sustainable Development: A Synthetic Meta-index Approach. Catholic University of Leuven and Wageningen University.
- Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., 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.
- Cherchye, L., Moesen, W., Van Puyenbroeck, T., 2004. Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *J. Common Market Stud.* 42 (5), 919–955.
- Chiang, C.-M., Lai, C.-M., 2002. A study on the comprehensive indicator of indoor environment assessment for occupants' health in Taiwan. *Build. Environ.* 37 (4), 387–392.
- Cooper, W.W., Seiford, L.M., Tone, K., 2000. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers, The Netherlands.
- Directorate for Science, Technology and Industry (DSTI), 2003. *Composite Indicators of Country Performance: A Critical Assessment*. Organisation for Economic Co-operation and Development.
- Eisenberg, D., 2004. The mixed effects of precipitation on traffic crashes. *Accident Anal. Prev.* 36, 637–647.
- European Transport Safety Council (ETSC), 2007. "Knowledge for Leadership"—The Road Safety Performance Index: Ranking Road Safety Performance Across the EU. Retrieved November 8, 2007, from <http://www.etsc.be/PIN.php>.
- European Transport Safety Council (ETSC), 2001. *Transport Safety Performance Indicators*.
- Expert Choice Inc., 2006. *Expert Choice 11*. Retrieved January 18, 2007, from <http://www.expertchoice.com>.
- Haas, R., Meixner, O., 2006. *An Illustrated Guide to the Analytic Hierarchy Process*. Presentation Institute of Marketing and Innovation, Vienna.
- Litman, T., 2007. Developing indicators for comprehensive and sustainable transport planning. In: In proceedings of the 86th annual meeting of the Transportation Research Board, Washington, DC.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2005. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Organisation for Economic Co-operation and Development.
- Odeck, J., 2005. Evaluating target achievements in the public sector: an application of a rare non-parametric DEA and Malmquist indices. *J. Appl. Econ.* 8, 171–190.
- Pedraja-Chaparro, F., Salinas-Jimenez, J., Smith, P., 1997. On the role of weight restrictions in data envelopment analysis. *J. Prod. Anal.* 8, 215–230.
- Pennoni, F., Tarantola, S., Latvala, A., 2005. *The 2005 European e-Business Readiness Index*. Joint Research Centre.
- Puolamaa, M., Kaplas, M., Reinikainen, T., 1996. *Index of Environmental Friendliness: a methodological study*. Eurostat/Statistics Finland.
- Saaty, T.L., 1980. *The Analytic Hierarchy Process*. McGraw-Hill, USA.
- SafetyNet, Work Package 3, 2005. *State-of-the-art Report on Road Safety Performance Indicators*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.
- Saisana, M., Tarantola, S., 2002. *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*. Joint Research Centre.
- Sharma, S., 1996. *Factor analysis*. In: *Applied Multivariate Techniques*. Wiley & Sons, USA.
- Sharpe, A., 2004. *Literature Review of Frameworks for Macro-indicators*. Centre for the Study of Living Standards.
- SWOV, 2007. *SafetyNet*. Retrieved November 8, 2007, from <http://www.erso.eu/safetynet/content/safetynet.htm>.
- Talbert, M.L., Balci, O., Nance, R.E., n.d., 1994. *Application of the Analytic Hierarchy Process to Complex System Design Evaluation*. Virginia Polytechnic Institute and State University.
- Tarantola, S., Liska, R., Saltelli, A., Leapman, N., Grant, C., 2004. *The Internal Market Index 2004*. Joint Research Centre.
- World Health Organization (WHO), 2004. *World Report on Road Traffic Injury Prevention*. Retrieved November 4, 2004, from http://www.who.int/world-health-day/2004/infomaterials/world_report/en.
- Yale Center for Environmental Law and Policy, Center for International Earth Science Information Network, 2005. *Environmental Sustainability Index*. Yale University and Columbia University.