A new approach for index construction: The case of the road user behavior index

Peer-reviewed author version

BABAEE, Seddigheh; Toloo, Mehdi; HERMANS, Elke & SHEN, Yongjun (2021) A new approach for index construction: The case of the road user behavior index. In: COMPUTERS & INDUSTRIAL ENGINEERING, 152 (Art N° 106993).

DOI: 10.1016/j.cie.2020.106993
Handle: http://hdl.handle.net/1942/33071

A New Approach for Index Construction: The Case of the Road User Behavior Index

Abstract

In recent years, composite indicators have become increasingly recognized as a useful tool for performance evaluation, benchmarking, and decision-making by summarizing complex and multidimensional issues. In this study, we focus on the application of data envelopment analysis (DEA) on index construction in the context of road safety and highlight the shortcomings of using the classical DEA models. The DEA method assigns a weight to each indicator by selecting the best set of weights for the unit under evaluation. The flexibility in selecting the weights in the classical DEA approach may lead to two interrelated problems: compensability and unfairness. These shortcomings are, respectively, overcome traditionally by imposing weight restrictions and applying a common weights approach. However, the problem of evaluating a layered hierarchy of indicators with a common set of weights (CSW) has not been addressed in the literature. To fill this gap, we propose a new approach for index construction to determine an optimal CSW to assess all units simultaneously while reflecting the hierarchical structure of the indicators in the model. The applicability of the suggested common-weight approach is illustrated by a case study on constructing a road user behavior index for a set of European countries. From a theoretical point of view, our approach provides a fair and identical basis for evaluation and comparison of countries in terms of driver's behaviors and, from a practical point of view, it significantly reduces the required computational burden for solving the formulated model. The obtained results clarify the sharper discrimination power of our model compared to the other methods in the literature.

Keywords: Road user behavior; Performance evaluation; Hierarchical structure; Composite indicators; Data envelopment analysis; Common set of weights.

1 Introduction

Road traffic crashes are the eighth highest cause of death globally, cutting short the lives of almost 1.35 million people every year (World Health Organization, 2018). Although road safety actions taken so far have been effective, the number of road fatalities and injuries is still unacceptably high. In order to improve road safety, one should consider not only the crash data but also the underlying risk factors influencing safety. In this respect, a road safety target hierarchy was proposed for the development of various indicators. The concept was first introduced in New Zealand's National Road Safety Strategy 2001–2010 (LTSA, 2000), in which road safety was represented as a pyramid consisting of four horizontal layers, from bottom to top respectively: safety measures and programs, safety performance indicators (SPIs), number killed and injured, and social cost. In the later European SUNflower study (Koornstra et. al, 2002), the layer "structure and culture" was added at the bottom of the pyramid to highlight the policy context of a country or its background conditions. The pyramid describes the relationships that exist between indicators in the road safety system indicating processes that lead to accidents and their consequent social costs (ETSC, 2001).

In Europe, based on the potential for different road safety domains to improve road safety, together with available data, the following seven risk factors have been identified as essential for road safety management and, hence, can be considered for the development of SPIs: drinking and driving, speeding, protective systems, daytime running lights, vehicles (passive safety), roads (infrastructure), and trauma management. These risk factors are in turn quantified by several appropriate SPIs that serve as the basis for assessing the current level of road safety in each country and checking progress, evaluating the impacts of various safety interventions, and comparing road traffic systems between countries and/or regions (Hakkert and Gitelman, 2007). Although individual indicators provide large amounts of information, it is not easy to compare countries by identifying an identical trend in many separate indicators because making a simple comparison for each indicator across countries only reveals a small (often trivial) portion of the road safety situation. As a result, in order to evaluate the multi-dimensional concept of road safety, it is beneficial to create a single composite indicator by consolidating a set of performance indicators.

In this study, we focus on road user behavior, which is recognized as the main contributor in the majority of all crashes (Stanton and Salmon, 2009). In a European SafetyNet project, Hakkert, Gitelman and Vis (2007) identified the most common risk factors for measuring the performance of road user behavior as drink driving, speeding, and nonuse of protective systems, each having various indicators. **Figure 1** depicts the *road user behavior* index, which is constituted from three behavioral factors: (i) drink driving, (ii) driving at inappropriate or excessive speed, and (iii) the use of protective systems. Drink driving is represented by two indicators: percentage of drivers with a blood alcohol level beyond the legal limit in roadside police tests (I_1), and the percentage of fatalities caused by accidents involving at least one driver affected by alcohol (I_2). Driving at inappropriate or excessive speed is evaluated by two indicators: the mean speed and the speed limit violations in free flow traffic, i.e. the percentage of vehicles exceeding the speed limit. Since the level of risk varies on different road types, the indicators for speed are further segregated into motorways, rural roads, and urban roads ($I_3 - I_8$). The last behavioral characteristic, protective systems, is quantified by three indicators: the daytime seat belt rate, which is obtained by the percentage of wearing in front (I_9), and rear (I_{10}) seats of passenger cars and vans under 3.5 tons, and the percentage of daytime usage of child restraints (I_{11}).

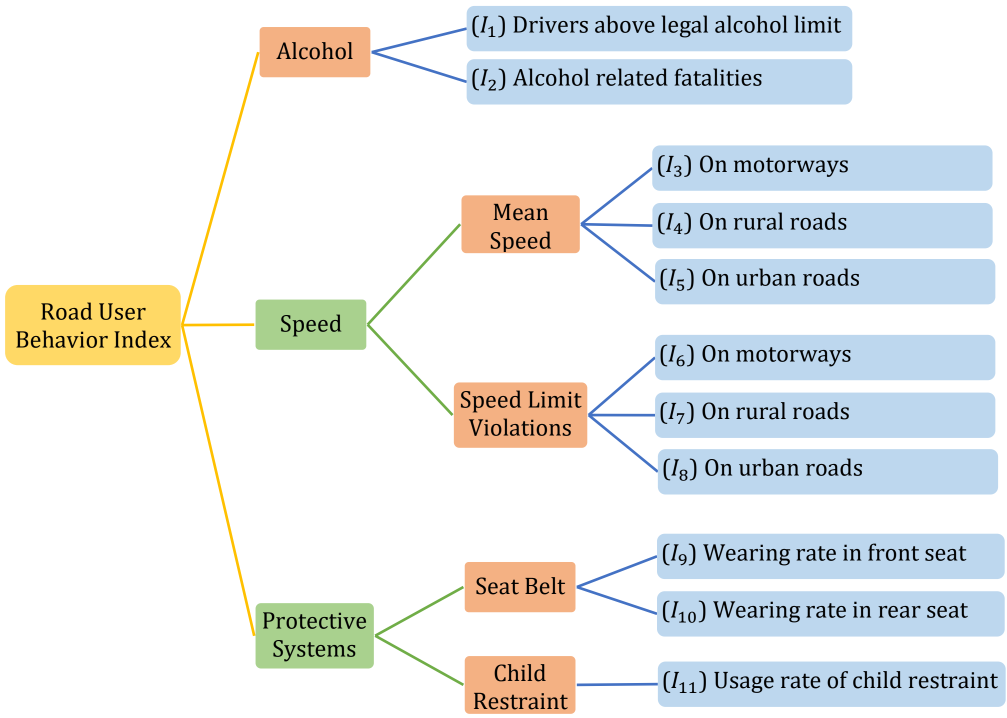


Figure 1. The hierarchical structure of road user behavior (source: Shen, Hermans, Brijs, Wets (2013))

In summary, there are eleven SPIs which are combined into a single hierarchical index with the aim of evaluating the performance of road user behavior for some European countries. The road user behavior index comprises of three pillars: Alcohol, Speed, and Protective systems. The first two indicators, i.e., I_1 and I_2, relate to Alcohol; the next six indicators relate to Speed via Mean Speed (I_3-I_5) and Speed Limit Violations (I_6-I_8) sub-indexes; the last three indicators, i.e., I_9-I_{11}, relate to Protective systems. For instance, increasing the seat belt wearing rate in the front and rear seats leads to an increase in the seat belt sub-index which enhances the protective systems pillar and subsequently improves the road safety situation in a country. Calculating the index requires a scientifically sound methodology that considers the layered hierarchy of the indicators and hence provides a sound basis for cross-country comparison.

The main approach for constructing an index is to assign a weight to each indicator and then aggregate them into a single index. Such aggregations and assignments directly affect the quality and reliability of the calculated index (see Greco, Ishizaka, Tasiou, Torrisi , 2019, Nardo et al., 2005, Saisana, Saltelli, Tarantola, 2005). Data envelopment analysis (DEA) has been accepted as a promising method that assigns the best set of weights to SPIs and then aggregates them into an index with a maximum possible score without relying on a priori knowledge on finding the weights. In fact, the relative performance of a specific country is calculated by considering the performance of all the other countries by solving a linear optimization problem. As a matter of fact, DEA finds the best possible weights directly from the data. The term 'best' means that the measured index score of each country is maximized relative to the others when these weights are assigned to the indicators. However, this flexibility in selecting the weights in the traditional DEA approach may lead to two interrelated issues: compensability and unfairness (Hatami-Marbini and Toloo, 2016). The former problem arises when an extreme (very low or very high, unrealistic) value is assigned to a weight, which is in conflicting contrast to decision-makers' beliefs. The latter problem refers to the situation in which variable weights are assigned to a single indicator rather than a fixed weight that deters the evaluation process from a fair and identical condition. To overcome these shortcomings, imposing suitable weight restrictions, respectively, applying the common-weights (CW) approach was developed within the DEA literature (for a deeper discussion we refer the reader to Hatefi and Torabi, 2010, Sun, Wu, Guo, 2013, and Wang, Luo, Lan, 2011). A point in common among them is that all indicators were simply treated as if they belong to the same layer. However, in creating an index, it is worthwhile to pay attention to the structure of the indicators, i.e., indicators that share similar conceptual features must be considered in the same category. Consequently, the indicators might belong to a different item and further be linked to one another creating a multilayer hierarchical structure, as shown in Figure 1. In this situation, considering the basic DEA models which consider a single layer for all the indicators results in ignoring the information available on their hierarchical structure.

To overcome this limitation, Meng, Zhang, Qi, Liu (2008) suggested a layered hierarchy DEA approach in which the weights among categories are determined by the DEA model and the weights within categories (internal weights) are determined by the weighted-average approach embedded in the DEA framework. However, their proposed model is non-linear and limited only to situations with a two-layer hierarchy. Later, Kao (2008) by using variable substitution, transformed it into a linear form. Thereafter, as a valuable extension, Shen et al. (2011, 2013) developed a generalized multi-layer DEA (MLDEA) and MLDEA based index (MLDEA-I) model, which consider the layered hierarchy of indicators, without any limitation on the number of layers. Nevertheless, the weights derived by these methods are not the same for all the DMUs which deters the evaluation process (i.e., ranking and benchmarking) from a fair and identical condition.

This study proposes a new approach for index construction with an optimal common set of weights (CSW) to evaluate all DMUs simultaneously whereas the hierarchical structure of the indicators is considered. Specifically, the concept of CSW is integrated into the model of Shen et al. (2013) leading to a new model with sharper discrimination power. In doing so, not only it contributes to the set of methods currently available for index construction, but also it proposes an extension of common weights to a case of hierarchically structured indicators. To the best of our knowledge, this paper is the first to combine the common weights approach with the hierarchical structure of indicators. The suggested approach leads to index values composed of similar weights which are important for a fair evaluation.

The rest of this paper is organized as follows: the next section presents the most important road safety indexes. Section 3 briefly reviews the classical DEA based index model and highlights some shortcomings of it. In Section 4, we propose a new method to generate an index with a CSW with considering the hierarchical structure of the indicators. The application of the proposed model is presented in Section 5, by a case study on benchmarking the road user behavior for a set of European countries. The paper ends with the conclusions in Section 6.

2 Road Safety Indexes

In the traditional approach, the safety performance of countries was assessed mainly by the safety outcomes in terms of fatalities per head of population, vehicle fleet, or exposure

(Gitelman, Doveh, Hakkert, 2010; ITF, 2017). Currently, benchmarking of road safety at the (inter)national level is being done using the concept of composite indicators in which road safety performance indicators are combined into an overall index. Prominent undertaken research studies on the road safety index (RSI) applied both objective and subjective approaches (See **Table 1**).

Table 1. Summary of the existing road safety indices

Ref.	The set of indicators	Index construction methodology	Aims
Al Haji, G. (2007)	Three pillars of the road safety domain: (i) Human performance which shows how safe the behavior of the road users is (ii) System performance comprised of safer roads, safer vehicles, enforcement, and socioeconomic performance (iii) Product performance considering fatality rates	- Equal weighting - Principal component analysis - Expert judgments - Weights based on previous experience	To assess the road safety performance at two groups of Highly motorized and less motorized countries
Hermans et al (2008)	- Alcohol and drugs - Speed - Protective systems - Vehicle - Infrastructure - Daytime running lights	-Equal weighting -Factor analysis -Analytic hierarchy process -Budget allocation -DEA	To evaluate the safety performance of 21 EU countries by applying 5 different weighting methods and selecting the best one for the context of road safety
Gitelman et al. (2010)	-Policy performance indicators -Final road safety outcome indicators -Intermediate outcome indicators -Background characteristics of the countries	-Factor analysis -Principal component analysis	To explore different trials in creating a road safety performance index for 27 European countries
Shen et al. (2011)	-Road user behavior indicators (i.e., Alcohol, speed, protective systems) as the model's input -Road safety final outcomes (# fatalities per million population, # serious injuries per million population, # slight injuries per million population, # crashes per million population) as the output	Multiple Layer Data Envelopment Analysis	To measure the road safety efficiency of thirteen European countries
Shen et al. (2013)	The same set of road user behavior indicators used in our study	Multiple Layer Data Envelopment Analysis based index	To assess the road safety performance of thirteen European countries
Bao et al. (2012)	Alcohol and drugs, speed, protective systems, vehicle, roads, and trauma management	Enhanced Fuzzy Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)	To assess the road safety performance of some European countries
Chen et al. (2015)	- Indirect dimension including human, vehicle, road, environment and management factors - Direct dimension: personal and traffic risk	Integrated entropy, TOPSIS, and Rank Sum Ratio methods	To assess the safety performance of 31 provinces in China
Chen et al. (2016)	-Policy performance indicators -Road safety performance indicators including final and intermediate outcome	Entropy-embedded RSR	To rank and classify some European countries and find the best-in-class as a benchmark in each group

For instance, Al-haji (2007), proposed a road safety development index to indicate the severity of the road safety situation in a set of countries by focusing on three pillars of the road safety domain: (i) Human performance which shows how safe the behavior of the road users is, (ii) System performance which is comprised of safer roads, safer vehicles, enforcement and socioeconomic performance, and (iii) Product performance considering fatality rates. He applied four weighting approaches including equal weighting, principle component analysis, expert judgments and weights based on previous experience, for the index construction.

In the SafetyNet project, Hakkert and Gitelman (2007) defined seven risk factors for road safety management. Thereafter, Hermans, Van den Bossche, and Wets (2008) suggested a safety performance indicator for each risk factor and investigated the assignment of different weighting methods to the individual indicators. The authors employed five commonly used approaches for the RSI creation, i.e., equal weighting, factor analysis, analytic hierarchy process, budget allocation, and DEA. Furthermore, the theoretical consideration, as well as the pros and cons of each method were described and the results were compared with the mortality rate as a relevant reference. Hermans, Van den Bossche, and Wets (2008) concluded that the DEA is the best approach for the RSI construction since it resulted in the highest correlation with the ranking of countries based on the mortality rate. Gitelman, Doveh, and Hakkert (2010) proposed an RSI by taking into account relevant indicators from different layers of the road safety pyramid. They utilized four categories of road safety indicators in their study: policy performance indicators, final road safety outcome indicators, intermediate outcome indicators, and the background characteristics of the countries (e.g. motorization level, population density). Two statistical weighting schemes, namely factor analysis and principal component analysis, were applied.

Contrary to the others who ignored the hierarchical structure of SPIs in the analysis, Shen (2012) conducted several outstanding studies on the consolidation of risk indicators considering the hierarchy of safety performance indicators. He identified six leading road safety risk indicators within three main road transport components to develop a comprehensive set of hierarchically structured safety performance indicators. Shen (2012) employed various extensions of the DEA method to construct an RSI for cross-country comparison (c.f. Shen et al., 2013).

Bao et al. (2012) utilized the well-known Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to assess the road safety performance of some European countries and proposed an enhanced hierarchical fuzzy TOPSIS model to integrate the multilayer SPIs and value judgment from experts in the form of linguistic expression into an overall index. Thereafter, Chen, Wang, and Deng (2015) integrated entropy, TOPSIS, and Rank Sum Ratio methods to assess the safety performance of 31 provinces in China.

Based on the resulting index scores, all the provinces were ranked and then categorized into five groups based on their level of road safety. Chen, Wu, Chen, Wang, and Wang (2016) also investigated a new approach, named Entropy-embedded RSR, to construct the RSI score with the aim of ranking and classifying some European countries and finding the best-in-class as the benchmark in each group. By comparing the results with other methods (e.g. SUNflower approach), the authors came to the conclusion that the proposed approach can facilitate comprehensive benchmarking of the countries. As there is no universally agreed-upon methodology to construct an index, it is worth exploring new approaches.

3 Data Envelopment Analysis (DEA)

DEA has been originated by Charnes, Cooper, Rhodes (1978) as a data-driven, nonparametric, optimization-based benchmarking technique that uses linear programming to evaluate the relative efficiency of a set of similar decision-making unit (DMU) with multiple inputs and multiple outputs. DEA has immediately been identified as a useful decision support system for performance evaluation, benchmarking, and decision-making and it has successfully been applied in a wide variety of research areas (Emrouznejad, Yang, 2017). Recently, Toloo, Mirbolouki (2019) extended an approach for selecting the best composite project using DEA. During the last two decades, significant attention has also been drawn to the application of DEA in index construction, i.e., integrating some indicators into a single composite index which is also known as the "benefit of the doubt" (BOD) approach (Cherchye, Moesen, Rogge, Puyenbroeck, 2007). A DEA based index (DEA-I) model is a conventional DEA model with multiple outputs without explicit inputs (for a deeper discussion of DEA-I models, we refer the readers to Liu, Zhang, Meng, Li, Xu, 2011, Toloo, 2013, and Toloo, Tavana, 2017). Assume that there are n DMUs ($\mathrm{DMU}_j; j = 1, \dots, n$) to evaluate in terms of s indicators ($\boldsymbol{y}_1, \dots, \mathbf{y}_s$), the DEA-I model for the DMU under evaluation, DMU_o, can be formulated as follows:

$$
\begin{aligned}
& I_o = \max \textstyle\sum_{r=1}^{s} u_{r_o} y_{ro} \\
& \text{s.t.} \\
& \textstyle\sum_{r=1}^{s} u_{r_o} y_{rj} \leq 1 \quad && j = 1, \dots, n \\
& u_{r_o} \geq 0 && r = 1, \dots, s
\end{aligned}
\tag{1}
$$

where u_{r_o} is the weight (a non-negative decision variable) for the r^{th} indicator $\boldsymbol{y}_r$ for $r = 1, \dots, s$ and the optimal objective value I_o is the composite indicator or the index score of

DMU_o. To get the highest or maximum index score, the above model assigns an optimal weight to each indicator in favor of DMU_o subject to a condition that the weighted sum of indicators for all DMUs does not exceed one. Model (1) is solved n times, one time for each DMU, and hence the optimal set of weights can differ from one DMU to another. Cooper, Seiford and Tone (2007) proved that the basic DEA models are always feasible and their optimal objective value is bounded, i.e., $0 < I_o \leq 1$. DEA partitions all the DMUs into two mutually exclusive and collectively exhaustive sets: efficient and inefficient DMUs. DMU_o is efficient if I_o = 1 and otherwise it is inefficient. Moreover, a higher index value implies a better overall relative performance.

Model (1) seeks the optimal weights in favor of the DMU under evaluation and this flexibility may lead to different weights for an indicator. Moreover, the optimal weights may not be unique for an efficient DMU. Such weight flexibility prevents to have a fair comparison with identical optimal weights for all the DMUs. To overcome this issue, Cook, Roll, Kazakov (1990) and Roll, Cook, Golany (1991) introduced the common-weights (CW) approach in the context of applying DEA to calculate the performance of highway maintenance patrols in Ontario.

Recently, more indicators which often have hierarchical structures are developed and applied in the construction of indexes, such as road safety performance indicators (see ETSC, 2001 and Hakkert, Gitelman, 2007). In this situation, it is not advisable to use the conventional DEA models anymore, because in these models it is assumed that all indicators are in a single layer. Shen et al. (2013) imposed some interesting weight restrictions on the DEA model to formulate an innovative MLDEA approach which is empowered to define indicators in different layers. The next section discusses a new CW Multi-layer DEA Index (CMI) approach, reducing the weight flexibility of Shen et al. (2013)'s model.

4 The Proposed Approach

The process of constructing a composite index with CSW in the presence of hierarchical layers comprises of five steps: Step 1, data gathering and processing; Step 2, imposing the set of weight restrictions to hold the given hierarchical structure of the indicators; Step 3, formulating a CMI model considering the provided weight restrictions which maximizes the performance of all DMUs, simultaneously; Step 4, calculating the index scores by multiplying the CSW obtained from the model with the (normalized) indicator values; Step 5, Computing the target values and drawing useful policy insights from the model output. **Figure 2** depicts the schematic diagram of the proposed methodology including five steps.

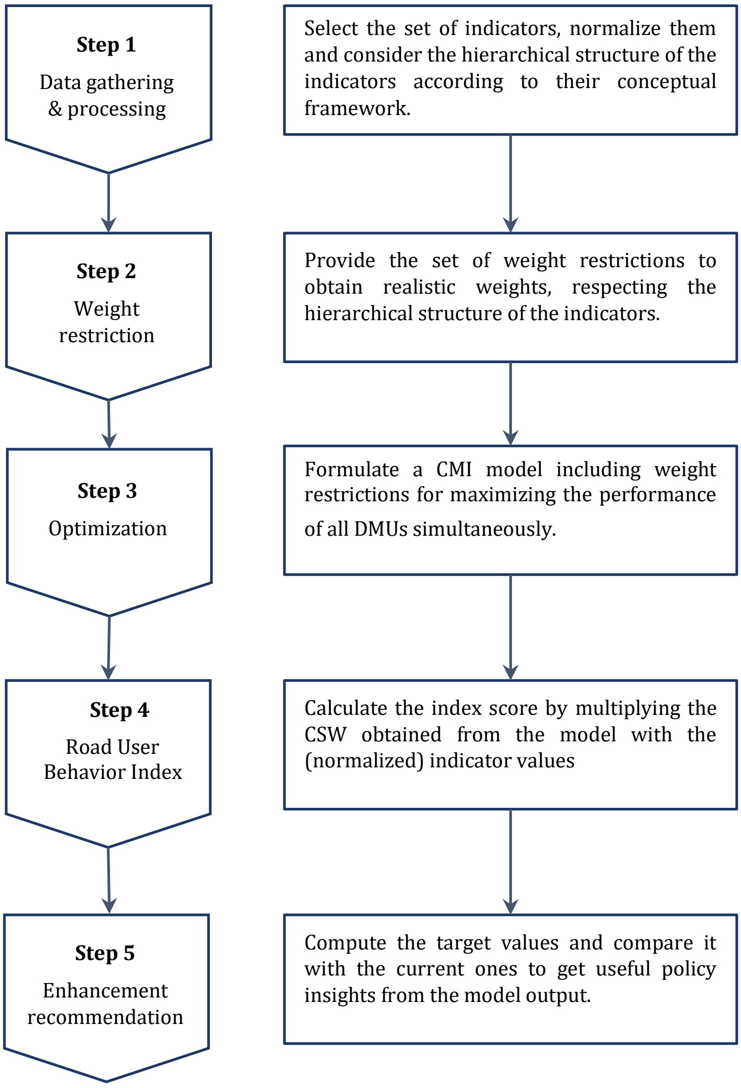


Figure 2. Schematic diagram of the proposed methodology

Assume that there are n DMUs to be evaluated with s indicators $\boldsymbol{y}_1, \dots, \boldsymbol{y}_s$ along with a K layered hierarchy, shown in **Figure 3**†. The first layer involves s indicators (nodes) $y_1^{(1)}, \dots, y_s^{(1)}$ which constitute $s^{(2)}$ indexes (nodes) in the second layer (in the same way,

† The figure can be interpreted as a tree with s leaves $y_1^{(1)}, \dots, y_s^{(1)}$ and a root DMU (graph theory).

the other layers are constructed). The value of each index (internal nodes) is measured as the normalized weighted sum of connected indicators (nodes) in the previous layer.

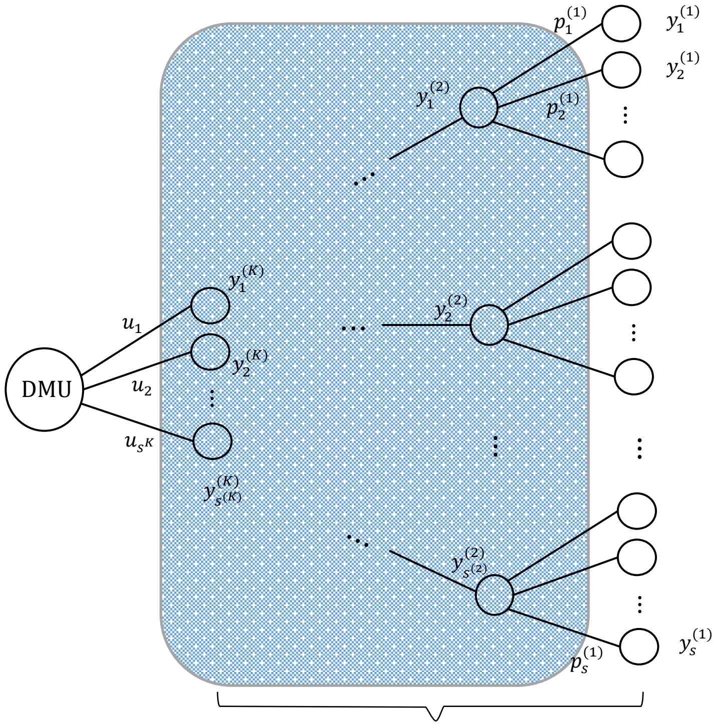


Figure 3. A typical hierarchical structure of indicators, adapted from Shen et al. (2013)

To clarify the concept of the common weight multi-layer DEA index construction, we continue explaining our proposed CMI model using a simple example which is depicted in **Figure 4**. It is assumed that $s^{(k)}$ is the number of categories in the k^{th} layer ($k = 1, \dots, K$). Note that with this definition in our example, we have $K = 3$ and $s^{(1)} = 11$; $s^{(2)} = 5$; $s^{(3)} = 2$. Moreover, let $A_{r_k}^{(k)}$and $p_{r_k}^{(k)}$ be the set of indicators and the internal weights belonging to the r^{th} category (node) in the k^{th} layer, respectively. The weighted sum of indicators for the DMU under evaluation in **Figure 4** is $\sum_{r_k=1}^{s^{(K)}} u_{r_k} y_{r_k}^{(K)}$ where $u_{r_k}$is the weight given to the r^{th} category in the K^{th} layer (i.e., the final layer).

According to our example, the weighted sum of indicators is equal to $u_1 y_1^{(2)} + u_2 y_1^{(3)} + u_3 y_2^{(3)}$. Considering the nodes of the last layer, we have $y_1^{(2)} = p_1^{(1)} y_1^{(1)} + p_2^{(1)} y_2^{(1)}$; $y_1^{(3)} = p_1^{(2)} y_2^{(2)} + p_2^{(2)} y_3^{(2)}$ and $y_2^{(3)} = p_3^{(2)} y_4^{(2)} + p_4^{(2)} y_5^{(2)}$. Therefore, the weighted sum of indicators for the DMU can be rewritten as $u_1\left(p_1^{(1)} y_1^{(1)} + p_2^{(1)} y_2^{(1)}\right) + u_2\left(p_1^{(2)} y_2^{(2)} +\right.$

$p_2^{(2)} y_3^{(2)}\Big) + u_3\Big(p_3^{(2)} y_4^{(2)} + p_4^{(2)} y_5^{(2)}\Big)$. In a similar manner, for a fixed node in the second layer we have $y_2^{(2)} = p_3^{(1)} y_3^{(1)} + p_4^{(1)} y_4^{(1)} + p_5^{(1)} y_5^{(1)}$ and so on. As a result, from a sequential substitution system, we can denote the objective function as:

$$\text{CMI} = \max \sum_{r_K=1}^{s^{(K)}} u_{r_K} \left(\sum_{r_{K-1} \in A_{r_K}^{(K)}} p_{r_{K-1}}^{(K-1)} \left(\cdots \sum_{r_2 \in A_{r_3}^{(3)}} p_{r_2}^{(2)} \left(\sum_{r_1 \in A_{r_2}^{(2)}} p_{r_1}^{(1)} \left(\sum_{j=1}^{n} y_{r_1 j} \right) \right) \right) \right) \quad (2)$$

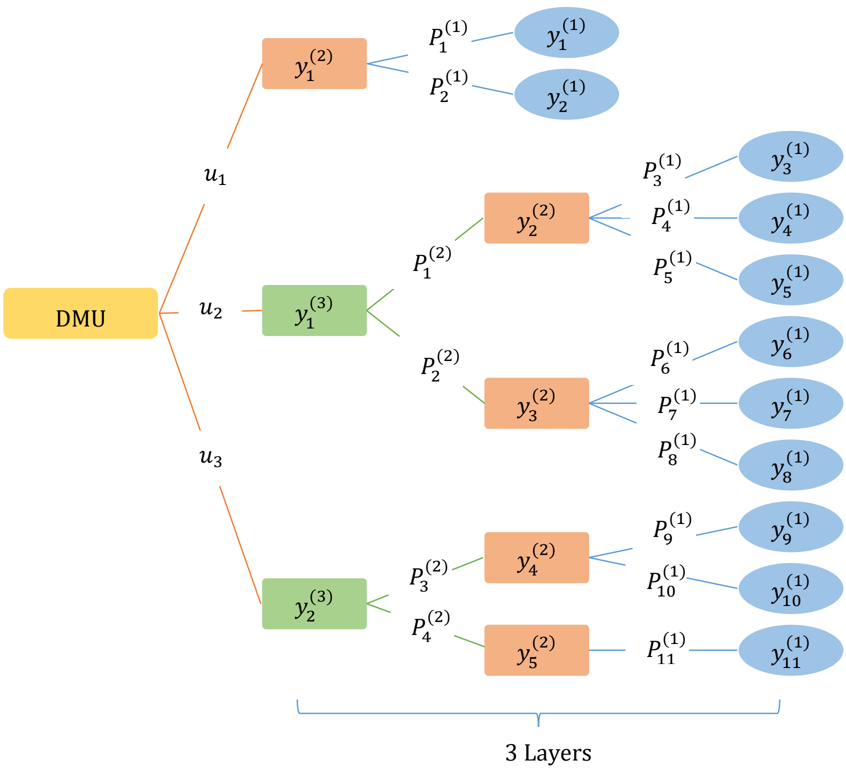


Figure 4. Three-layered hierarchical structure of indicators

which is corresponding to the following in our example:

$$\text{CMI} = \max\Big(u_1 p_1^{(1)} \left(\sum_{j=1}^{n} y_{1j}^{(1)}\right) + u_1 p_2^{(1)} \left(\sum_{j=1}^{n} y_{2j}^{(1)}\right) + u_2 p_1^{(2)} p_3^{(1)} \left(\sum_{j=1}^{n} y_{3j}^{(1)}\right) + u_2 p_1^{(2)} p_4^{(1)} \left(\sum_{j=1}^{n} y_{4j}^{(1)}\right) + u_2 p_1^{(2)} p_5^{(1)} \left(\sum_{j=1}^{n} y_{5j}^{(1)}\right) + u_2 p_2^{(2)} p_6^{(1)} \left(\sum_{j=1}^{n} y_{6j}^{(1)}\right) + u_2 p_2^{(2)} p_7^{(1)} \left(\sum_{j=1}^{n} y_{7j}^{(1)}\right) + u_2 p_2^{(2)} p_8^{(1)} \left(\sum_{j=1}^{n} y_{8j}^{(1)}\right) + u_3 p_3^{(2)} p_9^{(1)} \left(\sum_{j=1}^{n} y_{9j}^{(1)}\right) + u_3 p_3^{(2)} p_{10}^{(1)} \left(\sum_{j=1}^{n} y_{10j}^{(1)}\right) + u_3 p_4^{(2)} p_{11}^{(1)} \left(\sum_{j=1}^{n} y_{11j}^{(1)}\right)\Big)$$

This way, the existing hierarchical structure of the indicators is considered in the model and weights in one layer can be treated differently from the ones in another layer. However, since all the weights mentioned above are unknown (decision variables), their multiplication will turn the model nonlinear. More indicators in more layers extend the required computational time which makes the problem more complicated to deal with. Following Shen, Hermans, Brijs, and Wets (2012), we denote the weight for the indicators in the first layer by $\hat{u}_{r_k}$ which is measured as, i.e., $\hat{u}_{r_k} = u_{r_k} \prod_{k=1}^{K-1} p_{r_k}^{(k)}{}_{r_k \in A_{r_{k+1}}^{(k+1)}}$ for $r_k = 1, \dots, s$‡. In our example, we have:

$$\begin{aligned} \hat{u}_1 &= u_1 p_1^{(1)} & \hat{u}_6 &= u_2 p_2^{(2)} p_6^{(1)} \\ \hat{u}_2 &= u_1 p_2^{(1)} & \hat{u}_7 &= u_2 p_2^{(2)} p_7^{(1)} \\ \hat{u}_3 &= u_2 p_1^{(2)} p_3^{(1)} & \hat{u}_8 &= u_2 p_2^{(2)} p_8^{(1)} \\ \hat{u}_4 &= u_2 p_1^{(2)} p_4^{(1)} & \hat{u}_9 &= u_3 p_3^{(2)} p_9^{(1)} \\ \hat{u}_5 &= u_2 p_1^{(2)} p_5^{(1)} & \hat{u}_{10} &= u_3 p_3^{(2)} p_{10}^{(1)} \\ & & \hat{u}_{11} &= u_3 p_4^{(2)} p_{11}^{(1)} \end{aligned}$$

Thus, $\sum_{i=1}^{11} \hat{u}_i \left(\sum_{j=1}^{n} y_{ij}\right)$ should be maximized. As a result, we suggest the following CMI model with the aim of maximizing the performance of all DMUs simultaneously:

$$\begin{aligned} &\text{CMI} = \max \sum_{j=1}^{n} \left(\sum_{r_1=1}^{s} \hat{u}_{r_1} y_{r_1 j}\right) \\ &\text{s.t.} \\ &\sum_{r_1=1}^{s} \hat{u}_{r_1} y_{r_1 j} \le 1 \qquad && j = 1, \dots, n \\ &\hat{u}_{r_1} \ge 0 && r_1 = 1, \dots, s \end{aligned} \tag{3}$$

In addition, since the sum of the internal weights in each category of each layer is equal to one, i.e.,

$$\begin{aligned} &p_1^{(1)} + p_2^{(1)} = 1 && p_{11}^{(1)} = 1 \\ &p_3^{(1)} + p_4^{(1)} + p_5^{(1)} = 1 && p_1^{(2)} + p_2^{(2)} = 1 \\ &p_6^{(1)} + p_7^{(1)} + p_8^{(1)} = 1 && p_3^{(2)} + p_4^{(2)} = 1 \\ &p_9^{(1)} + p_{10}^{(1)} = 1 \end{aligned}$$

we have $u_1 = \hat{u}_1 + \hat{u}_2$; $u_2 = \hat{u}_3 + \hat{u}_4 + \hat{u}_5 + \hat{u}_6 + \hat{u}_7 + \hat{u}_8$ and $u_3 = \hat{u}_9 + \hat{u}_{10} + \hat{u}_{11}$. To illustrate the latter, we start from the right side of the above equation:

$$u_3 p_3^{(2)} p_9^{(1)} + u_3 p_3^{(2)} p_{10}^{(1)} + u_3 p_4^{(2)} p_{11}^{(1)} = u_3 \left[p_3^{(2)}\left(p_9^{(1)} + p_{10}^{(1)}\right) + p_4^{(2)} p_{11}^{(1)}\right]$$

‡ A tree has a unique path connecting each leaf to the root (see Bazaraa, Jarvis, Sherali (2010)). Let P be the path that connects indicator (leaf) r to the root. Then, $\hat{u}_r$ is measured as the product of the internal weights for each arc in P.

$$= u_3 p_3^{(2)} + u_3 p_4^{(2)} = u_3\left(p_3^{(2)} + p_4^{(2)}\right) = u_3$$

Analogously, $u_2 = \hat{u}_3 + \hat{u}_4 + \hat{u}_5 + \hat{u}_6 + \hat{u}_7 + \hat{u}_8$ and $u_3 = \hat{u}_9 + \hat{u}_{10} + \hat{u}_{11}$ which can be generalized as follows:

$$u_{r_K} = \sum\nolimits_{r_1 \in A_{r_K}^{(K)}} \hat{u}_{r_1} \tag{4}$$

Suppose that model (3) is solved and the common set of optimal weights, $\hat{u}_{r_1}^*$, is obtained. Thus, the internal weights of the indicators in each category of each layer can be realized as follows:

$$p_{r_k}^{(k)^*}{}_{r_k \in A_{r_{k+1}}^{(k+1)}} = \frac{\sum_{r_1 \in A_{r_k}^{(k)}} \hat{u}_{r_1}^*}{\sum_{r_1 \in A_{r_{k+1}}^{(k+1)}} \hat{u}_{r_1}^*} \quad r_k = 1, \dots, s^{(k)}, \ \ k = 1, \dots, K-1 \tag{5}$$

which is equivalent to the following in our example:

$$p_1^{(1)} = \frac{\hat{u}_1^*}{\hat{u}_1^* + \hat{u}_2^*} \ ; p_2^{(1)} = \frac{\hat{u}_2^*}{\hat{u}_1^* + \hat{u}_2^*}$$

$$p_3^{(1)} = \frac{\hat{u}_3^*}{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^*} \ ; p_4^{(1)} = \frac{\hat{u}_4^*}{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^*} \ ; p_5^{(1)} = \frac{\hat{u}_5^*}{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^*} \ ; \dots$$

$$p_1^{(2)} = \frac{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^*}{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^* + \hat{u}_6^* + \hat{u}_7^* + \hat{u}_8^*} \ ; p_2^{(2)} = \frac{\hat{u}_6^* + \hat{u}_7^* + \hat{u}_8^*}{\hat{u}_3^* + \hat{u}_4^* + \hat{u}_5^* + \hat{u}_6^* + \hat{u}_7^* + \hat{u}_8^*} \ ; \dots$$

The internal weights assigned to each indicator in each category of each layer indicates the importance level of the corresponding indicators. Therefore, to better reflect the reality of the situation, value judgment or experts' opinions can be integrated into the model by adding appropriate weight restrictions. To this aim, we can utilize various weight restriction techniques proposed in the DEA literature, such as absolute weight restrictions (Roll et al. 1991), assurance region or relative weight restrictions (Thompson, Singleton, Thrall, Smith 1986), ordinal weight restrictions and virtual weight restrictions (Wong, Beasley ,1990) (Cooper et al., 2007; Cherchye et al., 2007).

Now by incorporating the realized internal weights as additional constraints into model (3) and adding suitable weight restrictions, we obtain the following CMI model:

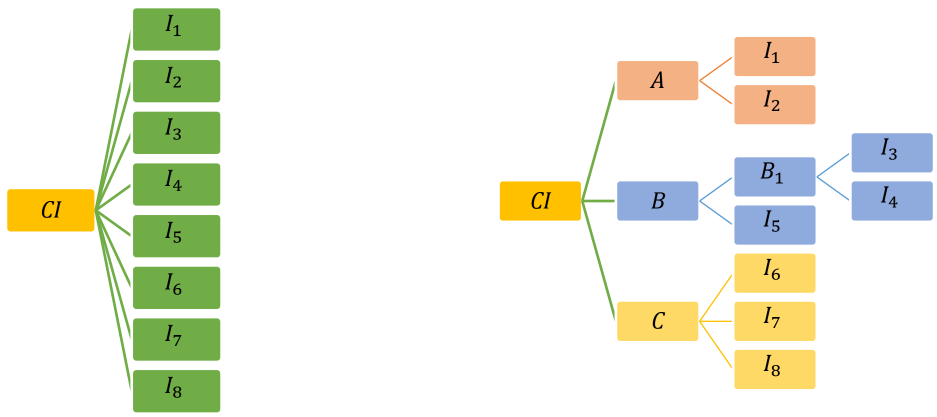
$$
\begin{aligned}
&\mathrm{CMI} = \max \textstyle\sum_{j=1}^{n}\left(\sum_{r_1=1}^{s} \hat{u}_{r_1}\, y_{r_1 j}\right) \\
&\text{s.t.} \\
&\textstyle\sum_{r_1=1}^{s} \hat{u}_{r_1}\, y_{r_1 j} \leq 1 && j = 1, \dots, n \\
&\frac{\sum_{r_1 \in A_{r_k}^{(k)}} \hat{u}_{r_1}}{\sum_{r_1 \in A_{r_{k+1}}^{(k+1)}} \hat{u}_{r_1}} = p_{r_k\ r_k \in A_{r_{k+1}}^{(k+1)}}^{(k)} \in \Theta && r_k = 1, \dots, s^{(k)}, k = 1, \dots, K-1 \\
&\hat{u}_{r_1} \geq 0 && r_1 = 1, \dots, s
\end{aligned}
\tag{6}
$$

where K stands for the number of layers in the layered hierarchy, $\hat{u}_{r_1}$ indicates the unknown weight for r_1^{th} indicators of the first layer ($r_1 = 1, \dots, s$), and Θ indicates the restrictions imposed on the corresponding internal weights. The objective function aims to maximize the index score of all DMUs simultaneously while satisfying the imposed weight restrictions. Moreover, let CMI_j be the composite indicator of DMU_j with the CSW, i.e., $\mathrm{CMI}_j = \sum_{r_1=1}^{s} \hat{u}_{r_1}\, y_{r_1 j}$. Then, mathematically $\mathrm{CMI} = \sum_{j=1}^{n} \mathrm{CMI}_j$. The first constraint set guarantees an intuitive interpretation of the composite indicator and implies that the composite indicator of no DMU is larger than one under the CSW. The second constraint set with the aim of reflecting the layered hierarchy of the indicators specifies the weights in each category of each layer and further restricts their flexibility.

It should be noted that although the suggested model (6) has some similarities with the classical model (1), it differs with respect to the following key features: the objective function and some additional constraints on the layer-specific weights. Before running the model, first, we have to aggregate the values of the indicators within each category of each layer by the weighted sum approach in which the sum of the internal weights is equal to one. Then, in the objective function, instead of using the indicator values for each DMU, and running the model n times, one time for each DMU separately, we calculate the sum of the indicator values of all DMUs and then run the model once, to get the optimal CSW for the indicators.

The suggested CMI approach has the following distinct advantages over the traditional DEA approach:

I. **Realistic**: the proposed model takes into account the hierarchical relationship that exists among the indicators, contrary to the traditional DEA models in which all the indicators are simply treated as if they belong to the same layer. Assume an index is going to be created with both the traditional and the proposed approach. As can be seen in **Figure 5**, the conceptual framework of the indicators and their hierarchy in the proposed approach (b) are kept, while in the traditional approach (a) all the indicators, i.e., I_1 *to* I_{11}, are considered in just one layer.



(a) Traditional approach (b) Proposed CW approach

Figure 5. Simplified versus realistic approaches

II. **One model to solve**: in the proposed model, the optimal set of weights is obtained by solving only a single integrated problem, in contrast to solving n problems in the traditional DEA models and there is no need to solve corresponding individual Linear Programming problems for evaluating all efficiencies/calculating all the index scores. In the proposed model, the index score of each DMU is calculated simply by multiplying the CSW obtained from model (6) with the (normalized) indicator values.

III. **Discrimination power**: the proposed model enables considering the hierarchical information of the indicators, and incorporating the weight restrictions in each category of each layer improves the discrimination power of the approach. Cooper et al. (2007) discussed that DEA models with restricted weight results possess sharper discriminating power. In other words, if a DMU is efficient in our model, then it is efficient in the traditional model. However, its reverse is not always true.

IV. **Reasonable and fair**: Our proposed approach provides a fair evaluation (ranking and benchmarking) since the performance of all the DMUs is evaluated based on the same set of weights under an identical condition. As mentioned earlier, traditional DEA models put each DMU in its best light by allowing each DMU to freely choose the weights of the indicators to maximize its performance (they are known to be optimistic). Therefore, solving the model for each DMU provides n different sets of weights for the indicators. The CW models solve a single optimization problem to obtain a set of weights that results in the highest overall efficiency of all DMUs (See **Figure 6**).

$$\begin{bmatrix} I_{11} & I_{12} & \dots & I_{1m} \\ I_{21} & I_{22} & \dots & I_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nm} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix}^T \Rightarrow \begin{bmatrix} CI_{11} & CI_{12} & \dots & CI_{1n} \\ CI_{21} & CI_{22} & \dots & CI_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ CI_{n1} & CI_{n2} & \cdots & CI_{nn} \end{bmatrix}$$

(a) Basic DEA approach

$$\begin{bmatrix} I_{11} & I_{12} & \dots & I_{1m} \\ I_{21} & I_{22} & \dots & I_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nm} \end{bmatrix} \begin{bmatrix} cw_1 \\ cw_2 \\ \vdots \\ cw_m \end{bmatrix} \Rightarrow \begin{bmatrix} CI_1 \\ CI_2 \\ \vdots \\ CI_n \end{bmatrix}$$

(b) Suggested CW approach

Figure 6. Graphical representation of CI construction in different approaches

V. **Target setting and enhancement recommendation:** The model output provides recommendations for inefficient DMUs on how to improve their performance. This feature is demonstrated in section 5 in the context of road safety.

5 Case study: Constructing a Road User Behavior Index

To validate the effectiveness of our proposed approach, we construct a road user behavior index for a set of European countries (or DMUs): Austria (AT), Belgium (BE), Finland (FI), France (FR), Hungary (HU), Ireland (IE), Lithuania (LT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SI), Sweden (SE), and Switzerland (CH) by applying the CMI model. The data of eleven hierarchical safety performance indicators for these countries have been adopted from Shen et al. (2013) and refer to the period 2006-2008. We use the same data set to compare the results of our proposed model with the MLDEA-I model. First, to tackle the different measurement units of the indicators and to ensure that all the indicators are expressed in the same direction with respect to their expected safety performance impact, the raw data have to be normalized. Among existing normalization methods (see OECD, 2008), the distance to a reference method (7) is used since the ratio of two numbers is best kept by this approach.

$$I_{rj} = \begin{cases} \dfrac{y_{rj}}{\max\{y_{rj}: j = 1, \dots, n\}} & \text{if } y_{rj} \text{ is a benefit} \\ \\ \dfrac{\min\{y_{rj} : j = 1, \dots, n\}}{y_{rj}} & \text{if } y_{rj} \text{ is a cost} \end{cases} \tag{7}$$

where I_{rj} and y_{rj} are the normalized and raw value of the individual indicator y_r, respectively, for DMU$_j$ ($j = 1, \dots, n$); "max y_{rj}" and "min y_{rj}" represent the maximum and minimum value of each indicator in the data set which are selected as the reference (or benchmark) for normalization when a benefit respectively a cost indicator is taken into account. Among the indicators, the eight SPIs related to alcohol and speed are undesirable indicators, while the three SPIs related to protective systems are desirable ones which means the higher the value of a given individual indicator, the better for the corresponding DMU. As a result, the DMU with the highest performance receives a normalized value of one whereas the others are expressed as the percentage share of that DMU's value. The resulting normalized data based on (7) are presented in **Table 2.**

Next, due to the hierarchical structure of the indicators, different preferences can be used at different levels, and as a result, the value judgment from decision makers or experts can be incorporated by restricting the weight flexibility in each category of each layer.

Table 2. Normalized data on the eleven hierarchical SPIs.

Country	Alcohol		Speed						Protective systems		
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}
AT	0.116	0.463	0.917	0.781	0.802	0.766	0.051	0.254	0.904	0.669	0.863
BE	0.068	0.654	0.827	0.743	0.768	0.348	0.029	0.222	0.799	0.467	0.729
FI	0.593	0.136	0.942	0.729	0.907	0.409	0.023	0.323	0.911	0.949	0.716
FR	0.263	0.123	0.912	0.787	0.838	0.505	0.037	0.318	1.000	0.957	0.937
HU	0.279	0.283	0.973	0.793	0.817	0.362	0.033	0.230	0.727	0.479	0.433
IE	0.237	0.119	0.924	0.762	0.724	1.000	0.032	0.223	0.901	0.875	0.857
LT	0.555	0.321	0.978	0.713	0.714	0.789	0.025	0.318	0.609	0.350	0.404
NL	0.081	1.000	0.879	0.740	0.881	0.454	0.020	0.234	0.959	0.852	0.758
PL	0.091	0.438	0.788	0.697	0.647	0.290	0.015	0.165	0.799	0.564	0.905
PT	0.137	0.610	0.828	0.618	0.919	0.302	0.014	0.360	0.881	0.549	0.591
SI	0.122	0.078	0.943	1.000	0.713	0.480	1.000	0.163	0.874	0.527	0.672
SE	1.000	0.357	0.864	0.717	0.870	0.241	0.019	0.259	0.973	0.887	1.000
CH	0.277	0.230	0.922	0.757	1.000	0.710	0.043	1.000	0.887	0.770	0.895

In this study, to be consistent with Shen et al. (2013), the same weight restrictions are imposed. i.e. to make sure that all the three risk factors – alcohol (Al.), speed (Sp.), and protective systems (P.S.) – will be represented to some extent in the overall road user behavior index, the share of each of these three risk factors is considered to lie within the range of [0.1,0.5], still wide enough to allow a high level of flexibility;

$$0.1 \leq Share_{Al.}, Share_{Sp.}, Share_{P.S.} \leq 0.5$$

It is worth noting that the share restitution is, in fact, the product of the indicator values and the corresponding weights, divided by the final index score:

$$0.1 \leq \frac{u_{r_K} \times y_{r_K}^{(K)}}{CMI} \leq 0.5, \qquad r_K = 1, 2, 3$$

Then for the rest, the SPIs belonging to the same category of each layer are considered to be of similar importance and are obligated to vary within a range of 0.8 to 1.2 of their average weights. It means that for the case of three indicators in a category, the weights are restricted to lie between $0.267\left(= 0.8 \times \frac{1}{3}\right)$ and $0.4\left(= 1.2 \times \frac{1}{3}\right)$ such as the three indicators with respect to the mean speed and the speed limit violations. Similarly, for the case of two indicators in a category, they are restricted to lie between $0.4\left(= 0.8 \times \frac{1}{2}\right)$ and $0.6\left(= 1.2 \times \frac{1}{2}\right)$ such as the two alcohol indicators, as well as the two seat belt indicators in the first layer, and the speed indicators and the protective systems indicators in the second layer. Now, the proposed CMI model can be applied to compute the optimal road user behavior index score of each country by combining 11 hierarchically structured indicators grouped into three different categories: alcohol, speed and protective systems. While in the MLDEA-I model each DMU obtains its own best possible indicator weights, the CMI model determines a set of weights to get the highest performance for all DMUs simultaneously, leading to a fair comparison among the DMUs. We have employed GAMS[§] optimization software to solve model (6).

[§] General Algebraic Modeling System, available at www.gams.com

Table 3 shows the best possible multiple layer common weight for each indicator, together with their rank in descending order on the third row.

Table 3. Optimal Multiple Layer CSW

Indicator	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}
CW	0.275	0.413	0.080	0.053	0.066	0.120	0.080	0.099	0.133	0.089	0.148
Rank	2	1	9	11	10	5	8	6	4	7	3

Based on the obtained optimal CW, the proportion of the three risk domains, i.e., alcohol, speed, and protective systems, and the percentage share of each indicator in the final index score are represented on the following pie chart, respectively.

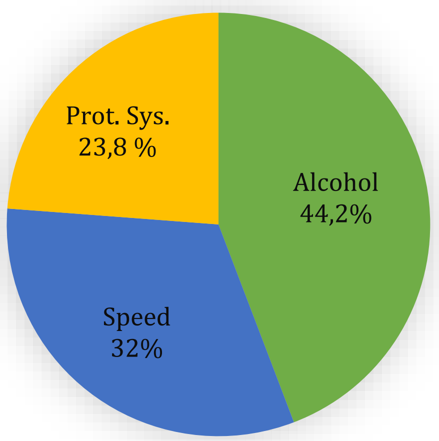


Figure 7. The proportion of the three risk domains in the final index score

Figure 7 indicates that alcohol with 44.2% accounts for the highest percentage share of the road user behavior index score followed by speed with 32% and protective systems with 23.8% in the third place.

Moreover, **Figure 8** illustrates the contribution of each indicator to the final index score. It shows that the percentage of alcohol-related fatalities (26.5%), the percentage of drivers above the alcohol limit (17.7%), the usage rate of child restraint (9.5%), the seat belt wearing rate in front seat (8.6%) and the percentage of vehicles exceeding the speed limit on motorways (7.7%) are the top five influential factors in the final index score.

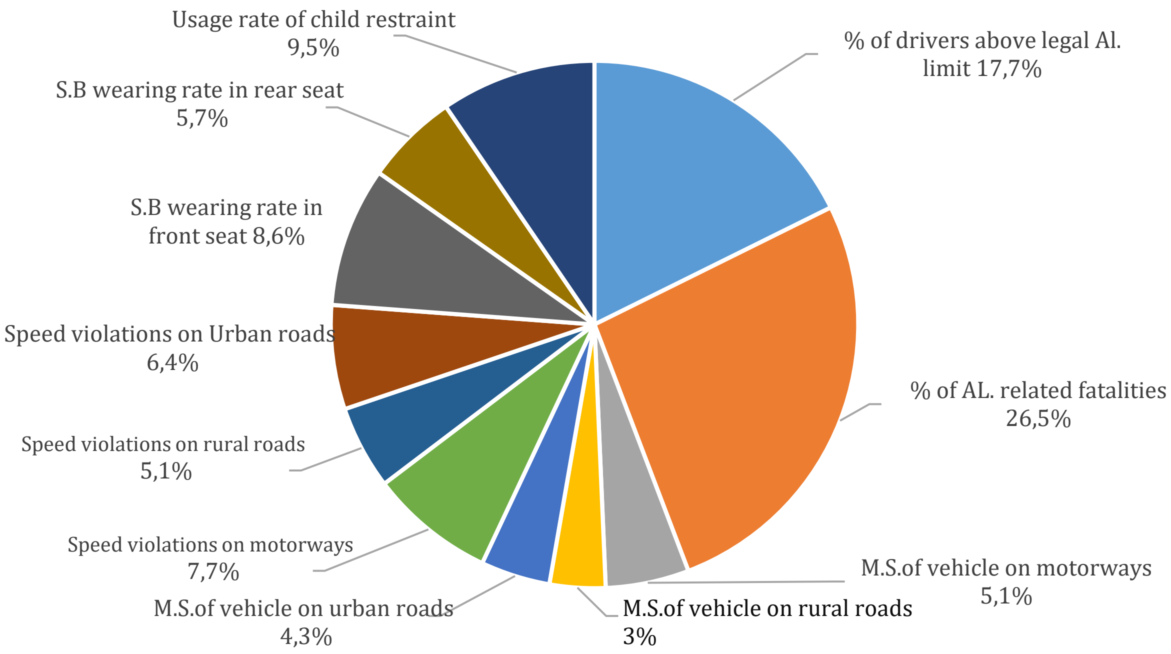


Figure 8. The percentage share of each indicator in the final index score

Having gained the CSW from model (6), the internal common weights for indicators in each category of each layer can be deduced using (5). For instance, consider $I_1 = 0.275$ and $I_2 = 0.413$ which belong to the category of alcohol. Their corresponding internal weight is equal to 0.400 $\left(= \frac{0.275}{0.275+0.413}\right)$ and 0.600 $\left(= \frac{0.413}{0.275+0.413}\right)$ respectively. (for more details, See **Figure 9**)

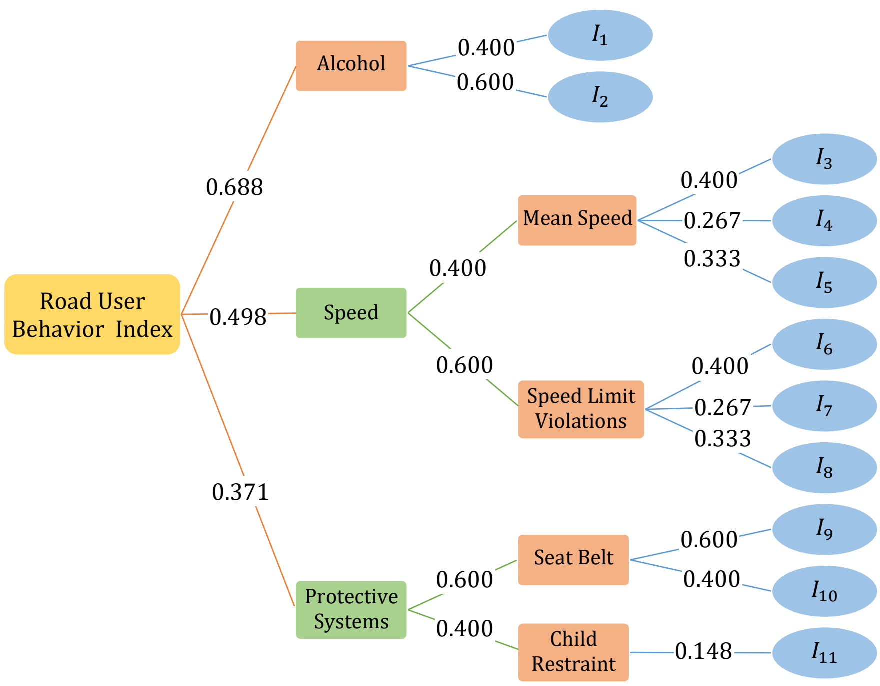


Figure 9. The internal Common Weights of hierarchically structured SPIs for all countries

Moreover, the index score of each country is calculated by multiplying the common weights obtained by solving model (6) with the normalized indicator values (**Table 2**). **Table 4** compares the road user behavior index score for the 13 European countries with different approaches. All index values lie between zero and one, with a value equal to one identifying the best performer, whereas a score of less than one indicates underperforming countries. The higher the index score, the better the overall relative performance in road user behavior.

Table 4. Road user behavior index scores for 13 European countries

Country	Basic DEA-I	Shen's model	New model
AT	**1.000**	0.935	0.820
BE	0.938	0.829	0.767
FI	**1.000**	0.925	0.788
FR	**1.000**	0.865	0.745
HU	**1.000**	0.728	0.640
IE	**1.000**	0.863	0.746
LT	**1.000**	0.846	0.748
NL	**1.000**	**1.000**	0.998
PL	0.955	0.806	0.692
PT	0.979	0.835	0.776
SI	**1.000**	0.679	0.658
SE	**1.000**	**1.000**	**1.000**
CH	**1.000**	**1.000**	0.858

By checking the index scores obtained from these models in **Table 4**, we see that the one layer model, i.e., the basic DEA-I model, was weak in distinguishing the best-performing countries from the underperforming ones, as most of the countries achieved an index score of one which is mainly due to the large number of indicators in comparison to the number of DMUs (or countries), and imposing the non-negativity as the only weight restriction into the model. In Shen's model, although by taking the layered hierarchy of indicators into account and incorporating weight restrictions in each category of each layer, the discrimination power of the model is improved, it is still not capable to discriminate between the following three countries: the Netherlands, Sweden and Switzerland. However, by applying the new model and generating the CW for indicators while considering the hierarchical structure of indicators and their corresponding weight restrictions, it can be seen that even with the same set of weight restrictions imposed to the models, the discriminative power of the proposed model is obviously improved and the optimal index score of one is obtained by only one country. It shows that Sweden is the best performing while the others are underperforming countries in terms of road user behavior.

Table 5 shows the ranking of the 13 European countries based on the values in **Table 4**. Furthermore, the number of road fatalities per million inhabitants which is traditionally used as a basis for countries' comparison, is also included in the analysis and is listed in the sixth column of **Table 5**. As can be seen, there is a relatively high degree of consistency between the index rankings across the different approaches, especially for those with better performing (such as Sweden, the Netherlands, and Switzerland) and with worst performing (such as Poland, Slovenia, and Hungary). The largest difference in the ranking is for Lithuania and France, each with five positions. Taking Lithuania as an example, it receives a rank of 8 based on the new model, but is ranked last out of 13 based on the number of fatalities per million inhabitants. To explain the reason for this wide deviation, first of all, we have to keep it in mind that the former considers a set of safety performance indicators, and consequently, the overall index-based country ranking is different from the traditional one that is only based on the fatality rate. Looking back to the indicator data, we find out that Lithuania suffers from very poor performance with respect to protective systems and speed limit violations on rural roads, nonetheless, it has relatively good scores on the other top influential indicators in the final index score, i.e. the alcohol related indicators. Accordingly, a better rank is obtained for Lithuania in the index-based ranking.

Table 5. The road user behavior index scores and ranking for 13 European countries

Country	New model	Ranking	Shen's model	Ranking	No. of fatalities*	Ranking
SE	**1.000**	1	**1.000**	1-3	47	1
NL	0.998	2	**1.000**	1-3	48	2
CH	0.858	3	**1.000**	1-3	49	3
AT	0.820	4	0.935	4	84	7
FI	0.788	5	0.925	5	67	4
PT	0.776	6	0.835	9	89	8
BE	0.767	7	0.829	10	96	9
LT	0.748	8	0.846	8	198	13
IE	0.746	9	0.863	7	75	6
FR	0.745	10	0.865	6	71	5
PL	0.692	11	0.806	11	142	12
SI	0.658	12	0.679	13	127	11
HU	0.640	13	0.728	12	117	10

* Average value of 2006-2008, source: OECD (2016)

To further explore the relationship among the three rankings and make a quantitative comparison, we use the correlation coefficient. According to **Table 6**, the ranking results derived from the new model and Shen's model are highly correlated with a correlation coefficient of 0.890. Concerning their comparison in terms of computational complexity, the latter is obtained by solving n individual LP problems, while the former is obtained by solving only one integrated LP problem for evaluating all DMUs. Moreover, compared with the ranking based on the number of road fatalities per million inhabitants, the index ranking derived from the new model and Shen's model are relatively close to each other, 0.758 versus 0.870.

Table 6. Pearson's Correlation Coefficients among different ranking results

	New model	No. of fatalities	Shen's model
New model	1.000	0.758	0.890
No. of fatalities		1.000	0.870
Shen's model			1.000

Target setting is the process of establishing a target for an inefficient country in which meeting the target makes the country efficient. We take Belgium as an example (See **Table 7**). To get useful policy insights from the model output, we first divide all the current indicator values by Belgium's index score to obtain the target values (see Cooper et al., 2007). The sum-product of the new indicator values and the CSW gained from our model gives us the new index score equal to one. It means that if Belgium reaches these target values, then it is an efficient country. Next, we compute the difference between the actual and target values for each indicator and finally rank them to prioritize the road safety risk domains that Belgium needs to give extra attention to increasing its relative performance. An indicator with a higher ranking score is more important than those with a lower ranking score. Hence, it indicates that mean speed on motorways and speed limit violations on rural roads (I_3 and I_7) are the most and the least urgent road safety aspect, respectively, for Belgium.

Table 7. Target setting based on model output for Belgium

Indicator	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	Index Score
Current Value	0.275	0.413	0.080	0.053	0.066	0.120	0.080	0.099	0.133	0.089	0.148	0.767
Target value	0.089	0.853	1.078	0.969	1.002	0.454	0.038	0.290	1.042	0.609	0.951	1.000
Difference	0.021	0.199	0.251	0.226	0.234	0.106	0.009	0.068	0.243	0.142	0.222	
Rank	10	6	1	4	3	8	11	9	2	7	5	

Analogously, we can rank all the indicators for inefficient countries, as exhibited in **Table 8** in which only risk domains with a growth rate higher than 12% have been considered. There is an exception only for the Netherlands where the growth rate is set to 0.2% since Sweden and the Netherlands are performing very close to each other in terms of road safety.

As can be seen, most countries should take policy actions on speed and protective systems, i.e., wearing a seat belt to enhance their road safety. It can be accomplished by enforcement and education. Switzerland should focus on speed mainly on urban roads and so on.

Table 8. The Road safety priorities for each country

Country	Alcohol		Mean Speed			Speed limit violation			Protective system		
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}
AT			1	5	4	6			2		3
BE		6	1	4	3				2		5
FI			2	5	4				3	1	6
FR			4	6	5	7			1	2	3
HU			1	3	2	7			4	5	6
IE			2	6	7	1			3	4	5
LT	6		1	4	3	2			5		
NL		1							2		
PL		7	3	4	5				2	6	1
PT		5	3	4	1				2		6
SI			3	1	5	8	1		4	7	6
CH					1			1			

Though countries can learn from actions taken in all other countries, from the practical point of view to set targets for improvement, it is more valuable to compare countries with the same level of motorization and safety development (Wegman et al., 2008). Therefore, it is of great importance to group comparable countries. In this study, using the hierarchical cluster analysis, and implementing the Average Linkage (between groups) method in SPSS 25, countries with similar safety levels based on the computed index scores are classified as follows (see **Figure 10** and **Table 9**):

Group 1: countries with a high level of road user behavior: Sweden and the Netherlands.

Group 2: countries with a moderate level of road user behavior: Switzerland, Austria, Finland, Portugal, Belgium, Lithuania, Ireland, and France.

Group 3: countries with a low level of road user behavior: Poland, Slovenia, and Hungary.

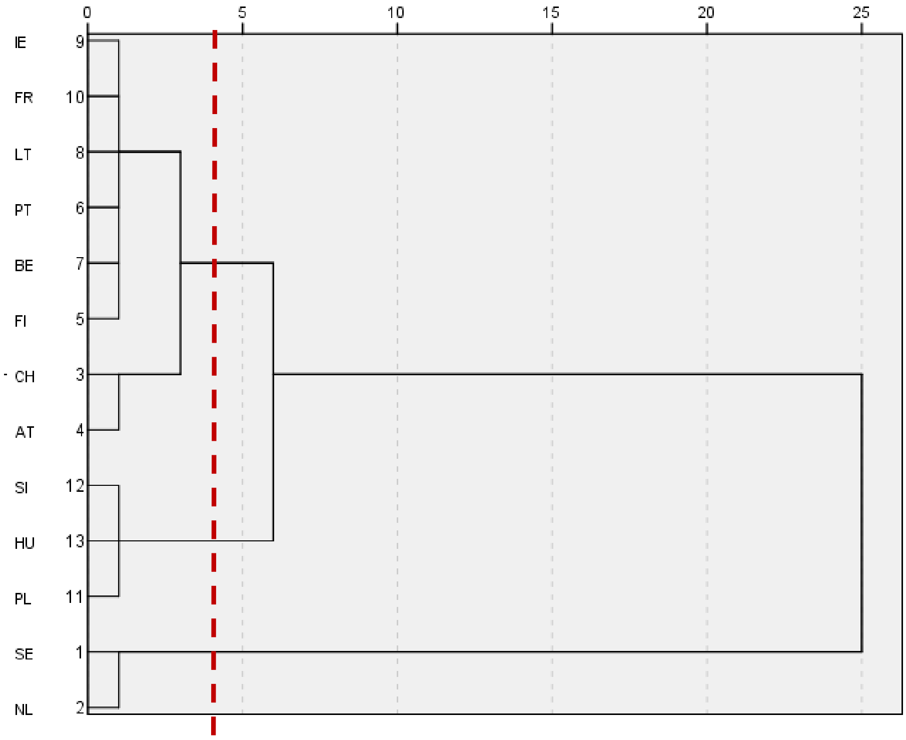


Figure 10. Dendrogram using the average linkage method

Consequently, the main characteristics of each group can be analyzed in detail to draw important conclusions on their features and provide recommendations for countries on how to improve their safety behavior**.

Table 9. Clustering countries based on the index scores using the Hierarchical Cluster Analysis

Group	Level	Country	Best country
I	High	SE, NL	SE
II	Moderate	CH, AT, FI, PT, BE, LT, IE, FR	CH
III	Low	PL, SI, HU	PL

** By performing the analysis of target setting and enhancement recommendations in each group

6 Discussion and Conclusion

In recent years, composite indicators have become increasingly recognized as a useful tool for performance evaluation, benchmarking, policy analysis, and public communication by summarizing complex and multidimensional issues in a relatively simple way. In this study, we focused on the application of DEA on index construction in the context of road safety and highlighted the shortcomings of using the classical DEA models. In particular, we dealt with two large methodological challenges: one was to reflect the hierarchical structure of the indicators into the model, and the other was to generate at the same time, the common weights for the indicators. To tackle the aforementioned issues, we proposed an integrated model with the aim of maximizing the performance of all DMUs simultaneously while reflecting the hierarchical structure of the indicators into the model. The advantages of the new approach have been identified and supported by the results of the case study on constructing a road user behavior index for a set of European countries. Using the 11 hierarchically structured SPIs related to road user behavior, the index scores were calculated for 13 European countries by selecting an optimal set of common weights under the imposed weight restrictions for the indicators, in each category of each layer. Thanks to the additional constraints on different layers, the flexibility of weights was reduced and consequently, the discrimination power of the proposed model was improved. Moreover, instead of solving the model n times, one model for each DMU as in the case of the classical DEA model, the proposed model aims, by solving a single optimization problem, to obtain a set of weights that results in the highest overall efficiency of all DMUs which greatly reduces the computational costs and provides a fair and identical platform for evaluation of DMUs.

Such a powerful tool is essential for benchmarking purposes: assessing the current level of road safety in each country, measuring the impacts of various safety interventions and monitoring the progress, and comparing road traffic systems between countries/jurisdictions thereby learning from the best practice policy. Nevertheless, benchmarking is not the end of the process and it needs to be followed by effective strategies, sufficient allocation of resources, successful implementation and persistent monitoring and evaluation in order to achieve continuous improvement over time. Though in this study we presented the application of the proposed model on benchmarking road user behavior for a set of European countries, it is transferable and feasible to apply it easily to a great number of applications in other areas.

Future work will require an uncertainty and sensitivity analysis of the index score. Since there exists a sequence of steps in the construction process of an index, it can be largely influenced by the methodological choices made during the index building process. Decisions concerning the selection of indicators, the normalization of the indicator values, the weighting of indicators and the way of aggregating, all can influence the final results. Therefore, it is of great importance to perform a sensitivity analysis to investigate the robustness of the index.

References

Al-haji, G. (2007). *Road Safety Development Index (RSDI): Theory , Philosophy and Practice*. *Science And Technology*. Linköping University.

Bao, Q., Ruan, D., Shen, Y., Hermans, E., Janssens, D. (2012). Improved hierarchical fuzzy TOPSIS for road safety performance evaluation. *Knowledge-Based Systems*, *32*, 84–90.

Bazaraa, M. S., Jarvis, J. J., Sherali, H. D. (2010). *Linear programming and network flows* (4th ed.). Hoboken, New Jersey: John Wiley & Sons.

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

Chen, F., Wang, J., Deng, Y. (2015). Road safety risk evaluation by means of improved entropy TOPSIS-RSR. *Safety Science*, *79*, 39–54.

Chen, F., Wu, J., Chen, X., Wang, J., Wang, D. (2016). Benchmarking road safety performance: Identifying a meaningful reference (best-in-class). *Accident Analysis and Prevention*, *86*, 76–89.

Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T. Van. (2007). An introduction to "benefit of the doubt" composite indicators. *Social Indicators Research*, *82*(1), 111–145.

Cook, W. D., Roll, Y., Kazakov, A. (1990). A DEA model for measuring the relative efficiency of highway maintenance patrols. *INFOR*, *28*(2), 113–24.

Cooper, W. W., Seiford, L. M., Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software* (2nd edi.). New York, Boston, Dordrecht, London, Moscow: Springer US.

Emrouznejad, A., Yang, G. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8.

ETSC. (2001). *Transport Safety Performance Indicators*. Brussels: European Transport Safety Council.

Gitelman, V., Doveh, E., Hakkert, S. (2010). Designing a composite indicator for road safety. *Safety Science*, *48*(9), 1212–1224.

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G. (2019). On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research*, *141*(1), 61–94.

Hakkert, A. S., Gitelman, V. (2007). *Road Safety Performance Indicators: Manual. Deliverable D3.8 of the EU FP6 project SafetyNet.* European Commission, Directorate-

General Transport and Energy.
Hatami-Marbini, A., Toloo, M. (2016). An Extended Multiple Criteria Data Envelopment Analysis Model. *Expert Systems with Applications*, *73*, 201–219.
Hatefi, S. M., Torabi, S. A. (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics*, *70*(1), 114–120.
Hermans, E., Van den Bossche, F., Wets, G. (2008). Combining road safety information in a performance index. *Accident Analysis and Prevention*, *40*(4), 1337–1344.
ITF. (2017). *Road Safety Annual Report 2017*. Paris: OECD Publishing.
Kao, C. (2008). A linear formulation of the two-level DEA model. *Omega*, *36*(6), 958–962.
Koornstra, M., Lynam, D., Nilsson, G., Noordzij, P., Pettersson, H., Wegman, F., Wouters, P. (2002). *SUNflower: a comparative study of the developments of road safety in Sweden, the United Kingdom, and the Netherlands*. *SWOV, Leidschendam*.
Liu, W. B., Zhang, D. Q., Meng, W., Li, X. X., Xu, F. (2011). A study of DEA models without explicit inputs. *Omega*, *39*(5), 472–480.
LTSA. (2000). *Road Safety Strategy 2010 : a consultation document.* Wellington, New Zealand.
Meng, W., Zhang, D., Qi, L., Liu, W. (2008). Two-level DEA approaches in research evaluation. *Omega*, *36*(6), 950–957.
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 Paper No. 2005/3). OECD Publishing.
OECD. (2008). *Handbook on constructing composite indicators*. *OECD Statistics Working Papers*.
Roll, Y., Cook, W. D., Golany, B. (1991). Controlling Factor Weights in Data Envelopment Analysis. *IIE Transactions*, *23*(1), 2–9.
Saisana, M., Saltelli, A., Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *168*(2), 307–323.
Shen, Y. (2012). *Inter-national Benchmarking of Road Safety Performance and Development using Indicators and Indexes. Data Envelopment Anamysis based Approaches.*
Shen, Y., Hermans, E., Brijs, T., Wets, G. (2013). Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model. *Social Indicators Research*, *114*(2), 739–756.
Shen, Y., Hermans, E., Brijs, T., Wets, G., Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident Analysis and Prevention*, *48*, 430–441.
Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T., Vanhoof, K. (2011). A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: A case study in road safety performance evaluation. *Expert Systems with Applications*, *38*(12), 15262–15272.
Stanton, N. A., Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, *47*(2), 227–237.
Sun, J., Wu, J., Guo, D. (2013). Performance ranking of units considering ideal and anti-ideal DMU with common weights. *Applied Mathematical Modelling*, *37*(9), 6301–6310.
Thompson, R. G., Singleton, F. D., Thrall, R. M., Smith, B. A. (1986). Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces*, *16*(6), 35–49.

Toloo, M. (2013). The most efficient unit without explicit inputs: An extended MILP-DEA model. *Measurement*, *46*(9), 3628–3634.
Toloo, M., Mirbolouki, M. (2019). A new project selection method using data envelopment analysis. *Computers and Industrial Engineering*, *138*, 106119.
Toloo, M., Tavana, M. (2017). A novel method for selecting a single efficient unit in data envelopment analysis without explicit inputs/outputs. *Annals of Operations Research*, *253*(1), 657–681.
Wang, Y.-M., Luo, Y., Lan, Y.-X. (2011). Common weights for fully ranking decision making units by regression analysis. *Expert Systems with Applications*, *38*(8), 9122–9128.
Wegman, F., Commandeur, J., Doveh, E., Eksler, V., Gitelman, V., Hakkert, S., … Oppe, S. (2008). *SUNflowerNext: Towards a Composite Road Safety Performance Index, Deliverable D6.16 of the EU FP6 project SafetyNet*. SWOV, Leidschendam.
Wong, Y.-H. B., Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, *41*(9), 829–835.
World Health Organization. (2018). *Global status report on road safety 2018: summary*. Geneva.