

18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015,
Delft, The Netherlands

Use of DEA and PROMETHEE II to assess the performance of older drivers

Seddigheh Babae^{a,*}, Maryam Bagherikahvarin^b, Renaud Sarrazin^c, Yongjun Shen^a,
Elke Hermans^a

^a Hasselt University, Transportation Research Institute (IMOB), 3590 Diepenbeek, Belgium

^b Université Libre de Bruxelles, 1050 Brussels, Belgium

^c MSM department, Belgian Road Research Center, 1200 Brussels, Belgium

Abstract

In recent years, there has been an increasing concern regarding the safety and mobility of elderly drivers. This study aims to evaluate the overall performance and ranking of a sample of 55 drivers, aged 70 and older, based on data from an assessment battery and a fixed-based driving simulator, by using the concept of composite indicators and multi criteria approach. To do so, drivers completed tests of an assessment battery of psychological and physical aspects as well as knowledge of road signs. Moreover, they took part in a driving simulator test in which scenarios that are known to be difficult for older drivers were included.

Composite indicators (CIs) are becoming increasingly recognized as a useful tool for performance evaluation, benchmarking and policy analysis by summarizing complex and multidimensional issues. One of the essential steps in the construction of composite indicators is aggregation and assignment of weights to each sub-indicator which directly affect the quality and reliability of the calculated CIs. In this regard, Data Envelopment Analysis (DEA) and Multi Criteria Decision Aiding (MCDA) have been acknowledged as two popular methods for weighting and aggregation and problem solving: ranking, sorting and choosing.

In this case study, on the one hand, we apply a DEA model to calculate the optimal performance index score for each driver. On the other hand, we apply a MCDA method to enrich the analysis of this problem by considering preferential information from Decision Makers (DM). This also results in a ranking of drivers in terms of driving performance.

The results of this study show that the best and the worst drivers identified by the two models are similar. These observations point out the interest of using PROMETHEE II (Preference Ranking Organization Method for Enrichment Evaluations) and DEA. The high correlation between these results confirms the robustness of our answers.

* Corresponding author. Tel.: +32(0)11 26 91 64; fax: +32(0)11 26 91 99

E-mail address: Seddigheh.babae@uhasselt.be

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Delft University of Technology

Keywords: Multiple Criteria Decision Aiding; PROMETHEE II; Data Envelopment Analysis; Composite Indicator; Older Drivers' Performance.

1. Introduction

The number of elderly drivers is increasing as a result of demographic changes (Mathieson et al., 2013). Although driving helps elderly to maintain their independence and autonomy, ageing is associated with a decline in sensory, motor and cognitive abilities which affects the ability to drive safely. To help elderly drivers to be aware of their own abilities and weaknesses and to regularly check their driving performance, there is an increasing need for developing a reliable assessment procedure to determine whether a person is still fit to drive.

The aim of this study is to provide a method to screen older drivers and to assess their relative performance, using data from an assessment battery and a fixed-based driving simulator. Within a performance improvement framework, performance evaluation plays a critical role in identifying weaknesses and planning goals for improvement. In this regard, composite indicators (CIs) are increasingly recognized as a valuable tool for performance evaluation, benchmarking and policy analysis by summarizing complex and multidimensional issues such as driving performance. The quality and reliability of the calculated CI is however affected by the weighting and aggregation of the indicator values (OECD 2008). In this respect, Data Envelopment Analysis (DEA) and Multi Criteria Decision Aiding (MCDA) have been considered as two popular methods for problem solving. Evidently there is a strong correspondence between the problems tackled by DEA and the ranking problems in multicriteria analysis (Roy, B., 1985). Indeed, inputs and outputs in DEA should be viewed as criteria in MCDA. Moreover, Decision Making Units (DMUs) are as alternatives. Through others, this similarity has been clearly pointed out in the works of Belton and Vickers (Belton and Vickers, 1993) and has led to the creation of special interest groups to study the interactions between DEA and different approaches of MCDA. Among them we can cite the Doyle and Green work illustrating DEA as an aid to MCDA (Doyle and Green, 1993). Yilmaz and Yurdusev (2011) used a DEA method as a tool to solve a MCDA problem. Shang and Sueyoshi (1995), Liu (2003) and Takamura and Tone (2003) worked on DEA and AHP.

In this case study, we apply a DEA model; this optimisation method is able to calculate the best performance index score for each driver, taking the hierarchical structure of indicators into account. Based on the results, the best performers, as benchmarks, are distinguished from underperforming ones, and all drivers are ranked by computing their cross index scores.

On the other hand, we apply a MCDA method to enrich the analysis of this problem by considering preferential information from Decision Makers (DM). The PROMETHEE II outranking method is used to generate a complete ranking of drivers by pair wise comparison of all the drivers under study. This comparison is done for the raw and normalized data to quantify to what extent the normalization of the evaluations is impacting the drivers' ranking. Consecutively, we compute the correlation between the results.

The rest of this paper is organized as follows: in the next section we will introduce the indicators and related data. Section 3 will detail data analysis based on DEA and PROMETHEE II. In section 4 we will summarize and discuss the results. This paper ends with conclusions (section 5).

2. Indicators and Data

2.1. Participants

Subjects aged 70 and older were recruited through the Geriatrics department of the Jessa Hospital with flyers distributed in the hospitals, senior associations and senior flats via local media. Participants had to hold a valid driver's license and still be active car drivers, with no stroke in the last four months and without any indication for dementia as assessed with the Amsterdam Dementia Screening (ADS) test. They had to have the physical ability to complete tests of a clinical assessment battery and simulator driving. 77 volunteers agreed to participate. Among

them, 22 participants were excluded due to simulator sickness. Thus, 55 participants remained in the sample (mean age = 76.49; standard deviation = 5.40).

2.2. Procedure

The test procedure consisted of two parts: First, a validated neuropsychological test battery including psychological and physical tests, as well as knowledge of road signs was administered at the Jessa Hospital. Next, a driving simulator test was conducted at the Transportation Research Institute of Hasselt University. For the purpose of this study, the following tests from the battery were incorporated in the analysis.

2.2.1. Psychological ability

The Mini Mental State Examination (MMSE)

The mini mental state examination (MMSE) is the most commonly used test for screening cognitive functioning. It is an 11-questions measure that investigates different areas of cognitive function: orientation to time and place, short term memory, registration (immediate memory), recall, constructional ability as well as language functioning (Folstein et al. 1975). The possible score ranges from 0 to 30. Scores of 25-30 out of 30 are considered normal. The higher the score, the better the psychological ability.

The Digit Span Forward (DSF)

DSF is originally part of the Digit Span Subtest of the Wechsler Adult Intelligence Scale (Wechsler 1955) where a random sequence of numbers is read by the examiner and the examinee recalls the numbers back. It assesses attention and working memory, as well as short-term verbal memory (Clark et al. 2011). Scores on this task are divided into four categories (0 = impaired, 1 = beneath average, 2 = average, 3 = above average). The more numbers a person can repeat correctly, the better the psychological ability.

Useful Field Of View (UFOV)

It is a computer-based test of functional vision and visual attention, which consists of three subtests measuring visual processing speed (UFOV 1), divided attention (UFOV 2), and selective attention (UFOV 3) (Edwards et al. 2005). It is recommended for use as a screening measure in conjunction with a clinical examination of cognitive functioning or fitness to drive. Scores are expressed in milliseconds for each subtest and range from 16.7ms to 500ms. Lower scores correspond with improved visual attention.

2.2.2. Physical ability

The Snellen Chart

This test is used for measuring visual acuity and is one of the most common clinical measurements of visual functioning (Rosser et al. 2001). Participants have to stand 6m from the whiteboard with several lines of black letters and read the lines. The more lines a person can read, the better the visual acuity (maximum score = 1.2).

The Get-Up-and-Go test

The Get-Up-and-Go test, also known as Timed Up-and-Go or Rapid Pace Walk, assesses mobility and balance of older adults (Carr et al., 2010). It measures, in seconds, the time taken by an individual to stand up from a standard arm chair, walk a distance of 3m, turn around, return and sit down again (Clark et al. 2011). Scores on this task are divided into three categories (0 = more than 20 seconds, 1 = between 11 and 20 seconds, 2 = less than 11 seconds). The faster one can complete the task, the better the motor ability.

The Four-test Balance Scale

This test is also used to assess motor abilities; more specifically, lower limb muscle strength and balance, with a maximum score of 1. An individual has to stand on 4 different foot positions of increasing difficulty - standing feet together, standing semi-tandem, standing tandem and one leg standing - for at least 10 seconds without an assistive device (Gardner et al. 2001).

2.2.3. Knowledge of Road Signs

The Road Sign Recognition (RSR) test is used to measure the knowledge of participants regarding road signs with a maximum score of 12 (Lundberg et al. 2003).

2.2.4. Driving ability evaluation

Driving performance was measured in a fixed-based medium-fidelity driving simulator (STISIM M400; Systems Technology Incorporated) with a force-feedback steering wheel, brake pedal, and accelerator. The visual virtual environment was presented on a large 180° field of view seamless curved screen, with rear view and side-view mirror images. Three projectors offer a resolution of 1024 × 768 pixels on each screen and a 60 Hz refresh rate. Data were collected at frame rate.

A 10 min practice session preceding the evaluation was implemented to allow participants to become familiar with the driving simulator. Information on the crash types of older drivers was taken into account and situations were included in the scenarios that are known to be difficult for older drivers. These situations are well documented in the literature. For instance, older drivers are over-represented in crashes occurring while turning off at intersections, where typically the older driver turns against oncoming traffic with right of way on the main road (Hakamies-Blomqvist, 1993; Zhang et al., 1998), gap acceptance while turning left at an intersection (Langford and Koppel, 2006; Yan et al., 2007) and response to signs, signals and road hazards (Horswill et al., 2010). A detailed description of the driving scenario is mentioned in Cuenen et al. (2015). The rides took place at inner-city (50 km/h) sections, outer-city (70-90 km/h) sections and highway (120 km/h) sections, in daylight and good weather conditions. The speed limit was indicated by the appropriate sign at the start of each outer-city and inner-city segment and repeated 30 meters after each intersection.

A total of 3 driving measures or indicators are used in the analysis: 1) Mean-Complete Stop which is computed from 200 meters before reaching the stop sign until the location of the stop. 2) Average Following Distance, between the driver and a lead vehicle in a road with a speed of 50 km/h and 70 km/h. 3) Mean driving Speed which is measured across separate road segments (i.e., 4.8 km) without any events (Triik et al., 2010) with the posted speed limits of 50, 70, 90 respectively 120 km/h.

3. Data analysis

In this study, to measure the multi-dimensional concept of driving performance, a composite indicator is created with respect to all aforementioned indicators for older drivers (visualized in Fig. 1). Simplistically, the composite indicator synthesizes the information included in the selected set of indicators in one score (Nardo et al., 2005). It should be mentioned that the driving measures and road sign recognition are considered in the group of driving ability.

Before the CI construction, normalization is carried out to tackle the different measurement units of the indicators. Among existing methods (Freudenberg 2003), the distance to a reference approach (OECD 2008) is used in this study since the ratio of two numbers is best kept by this approach. Thereafter, for the older-driver performance index construction, DEA is first applied. In doing so, the multiple layer model is adopted to take the hierarchical structure of the indicators into account and the cross index method is used for the ranking of the drivers. On the other hand, a MCDA method is applied to enrich the analysis of this problem by considering preferential information from DMs. PROMETHEE II (Preference Ranking Organization METHod for Enrichment of Evaluations) outranking method is used to generate a complete ranking of drivers by pair wise comparison of all the drivers under study. Both methodologies are explained in the following sections.

3.1. DEA for CI construction

3.1.1. Model

Data Envelopment Analysis (Cooper et al. 2007) is one of the most commonly used techniques for performance evaluation. It is a non-parametric optimization technique using a linear programming tool to measure the relative efficiency of a set of Decision Making Units (DMUs), or drivers in our study. Recently, there has been an increasing interest to the application of DEA in the construction of CIs (Despotis 2005; Cherchye et al. 2008; Hermans et al. 2008). By solving a linear programming problem, the best possible indicator weights are determined, and an optimal index score between zero and one is obtained for each unit, with a higher value indicating a better relative performance.

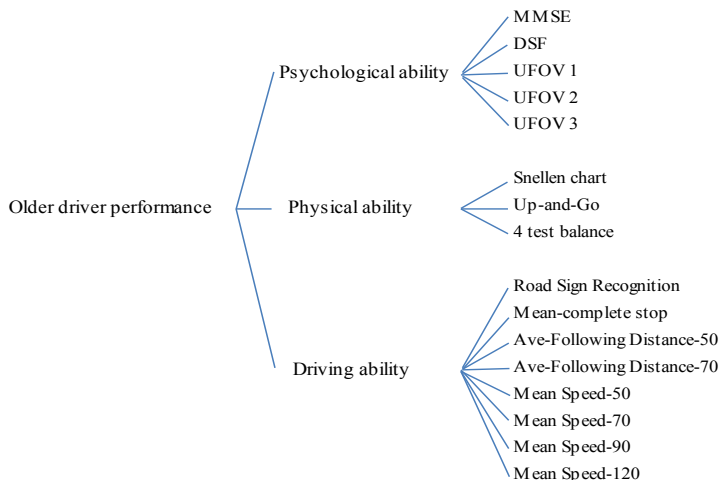


Fig. 1. Hierarchical structure of older driver's performance indicators.

In this study, to evaluate the driving performance of each older driver by combining all the 16 hierarchically structured indicators in one index score, a multiple layer DEA based composite indicator model (MLDEA-CI) developed by Shen et al. (2011, 2012) is adopted. The main idea of this model is to first aggregate the values of the indicators within a particular category of a particular layer by the weighted sum approach in which the sum of the internal weights equals one. Then, for the first layer, the weights for all the sub-indexes are determined using the basic DEA approach. More specifically, suppose that a set of n DMUs is to be evaluated in terms of s indicators (y) with a K layered hierarchy, the MLDEA-CI model can be formulated as follows (Shen et al. 2012):

$$\begin{aligned}
 CI_0 &= \max \sum_{f_1=1}^s \hat{u}_{f_1} y_{f_1 0} \\
 \text{s.t.} \quad & \sum_{f_1=1}^s \hat{u}_{f_1} y_{f_1 j} \leq 1, \quad j = 1, L, n \\
 & \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} / \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = w_{f_k}^{(k)} \in \Theta, \quad f_k = 1, L, s^{(k)}, \quad k = 1, L, K-1 \\
 & \hat{u}_{f_1} \geq 0, \quad f_1 = 1, L, s
 \end{aligned} \tag{1}$$

where

\hat{u}_{f_1} is the set of most favorable optimal weights for DMU₀, which are obtained by solving the model.

$s^{(k)}$ is the number of categories in the k^{th} layer ($k=1, 2, \dots, K$). $s^{(1)}=s$.

$A_{f_k}^{(k)}$ denotes the set of indicators of the f^{th} category in the k^{th} layer.

$W_{f_k}^{(k)}$ represents the internal weights associated with the indicators of the f^{th} category in the k^{th} layer, which sum up to one within a particular category and Θ indicates the restrictions imposed to the corresponding internal weights.

The subscript, o , refers to the driver whose index score is to be obtained by solving the constrained optimization problem, which maximizes the index value of the driver and satisfies the imposed restrictions. The first restriction guarantees an intuitive interpretation of the composite indicator and implies that no driver in the data set can be assigned an index value larger than one under these weights. With respect to the second restriction, the layered hierarchy of the indicators is reflected by specifying the weights in each category of each layer and further restricting their flexibility. In doing so, obtainment of realistic and acceptable weights is guaranteed.

In this study, to make sure that all the three aspects of a driver's abilities - psychological, physical and driving abilities - will be represented to some extent in the overall driving performance index score, each of these three factors is considered to have a similar importance in the final index score but still with 30% variability to allow a high level of flexibility in weight allocation. In addition, by the third restriction, all weights are restricted to be non-negative.

3.1.2. Cross index score

In addition, to fully rank all the drivers, the cross index method is employed. The main idea of this method is to evaluate the performance of a DMU using not only its own optimal weights, but also the ones of all other DMUs (Sexton et al. 1986). It means that a cross index matrix is to be developed in a way that the element in the i^{th} row and j^{th} column represents the index score of DMU j using the optimal weights of DMU i . Therefore, elements located on the diagonal are basic DEA indices. To rank the DMUs using the cross index method, the average of each column is calculated to obtain the mean cross index score.

3.2. PROMETHEE II

3.2.1. Model

Multiple Criteria Decision Aid (MCDA) techniques like MAUT (Keeney et al. 1979), AHP (Saaty 1980), ELECTRE (Roy 1991) and PROMETHEE (Brans, 1982) have been developed during the last five decades. Their objective is supporting Decision Makers (DM) in the selection of most compromise solution(s) and the ranking or sorting of alternatives. In this work, we focus on PROMETHEE II. The family of PROMETHEE methods is known due to their simplicity, number of applications in different fields such as finance, business, education, health care centers, insurance companies, etc. (Behzadian et al. 2010) and the existence of user friendly software, such as D-sight (Hayez, Q. et al., 2012).

The PROMETHEE method has been developed by J. P. Brans in 1982 and is based on pairwise comparisons. PROMETHEE II allows a DM to rank a finite set of n actions (DMUs in DEA) $A = \{a_1, \dots, a_j, \dots, a_n\}$ that are evaluated over a set of q criteria (like indicators in DEA) $F = \{f_1(a_j), \dots, f_k(a_j), \dots, f_q(a_j)\}$. Let $f_k(a_j)$ denote the evaluation of action a_j on criterion f_k . In what follows, we assume without loss of generality that criteria have to be maximized. First, the differences between every pair of actions on all criteria are computed as follows:

$$d_k(a_i, a_j) = f_k(a_i) - f_k(a_j), \forall a_i, a_j \in A, \forall k = 1, \dots, q \quad (2)$$

PROMETHEE is based on preference functions to integrate intra-criterion information. Thus in the second step, a generalized criterion $\{f_k(\cdot), P_k(a_i, a_j)\}$ is associated to each criterion. $P_k(a_i, a_j)$ represents the preference strength of action a_i over a_j . It is assumed to be a positive non-decreasing function of $d_k(a_i, a_j)$. The concept of preference function is used to transform the difference into a unicriterion preference degree; hence:

$$\pi_k(a_i, a_j) = P_k[d_k(a_i, a_j)] \quad (3)$$

The method provides the DM with a set of predefined preference functions for which at most two parameters have to be defined (the indifference and preference thresholds). The details of preference functions are explained in Brans and Mareschal (2002). The global preference degree between a_i and a_j , which varies between 0 and 1, is computed as follows:

$$\pi(a_i, a_j) = \sum_{k=1}^q \pi_k(a_i, a_j) \cdot w_k \quad (4)$$

Where $\pi(a_i, a_j) \geq 0$, $\pi(a_i, a_j) + \pi(a_j, a_i) \leq 1$ and w_k ($k = 1, \dots, q$) are normalized positive weights associated to the criteria. The positive and negative outranking flows are defined as follows:

$$\begin{aligned} \phi^+(a_j) &= \frac{1}{n-1} \sum_{x \in A} \pi(a_j, x) \\ \phi^-(a_j) &= \frac{1}{n-1} \sum_{x \in A} \pi(x, a_j) \end{aligned} \quad (5)$$

A complete pre-order, called PROMETHEE II can be obtained on the basis of the net flow score:

$$\phi(a_j) = \phi^+(a_j) - \phi^-(a_j) \quad (6)$$

Let us stress that the net flow score can also be computed as follows:

$$\phi(a_j) = \sum_{k=1}^q \phi_k(a_j) \cdot w_k \quad (7)$$

Such that:

$$\phi_k(a_j) = \frac{1}{n-1} \sum_{x \in A} \{\pi_k(a_j, x) - \pi_k(x, a_j)\}, \quad k = 1, \dots, q \quad (8)$$

The quantity ϕ_k is called the unicriterion net flow score of action a_j and is such that $-1 \leq \phi_k(a_j) \leq 1$. At this point, it is worth noting that the multicriteria problem can be viewed as an evaluation table (and associated parameters) or a matrix $\phi = (\phi_k(a_j))$. These values already integrate intra-criterion parameters and are all lying in the same range. In the next section we discuss the results from both methods.

3.2.2. Preference parameters in PROMETHEE model

In order to limit the complexity of the model and to respect the nature of each criterion, we use the usual preference function for most of the criteria in the problem. In this case, there is no need to choose the indifference and preference thresholds (q, p). Further, we apply the V-shape preference function for the criteria ‘‘UFOVtotal’’ ($q=0, p=2$) and ‘‘Snellen Chart’’ ($q=0, p=0.5$). We use the U-shape preference function for the criterion ‘‘RSR’’ ($q=p=2$).

The details in choosing preference functions and their thresholds are discussed in Brans and Mareschal (2002). All the criteria have to be maximized, except ‘‘UFOVtotal’’ which should be minimized. This criterion is an aggregation of the three UFOV criteria mentioned in section 2.2.1. We define equal weights for the three categories of criteria introduced previously (i.e. Psychological Ability, Physical Ability, Driving Performance). Then, we allocate equal positive weights to each sub-criterion of the category.

4. Results

In order to assess the robustness of the obtained results with the DEA model, we have modelled the problem with PROMETHEE II. We applied two different approaches in order to guarantee the independency of our results with the modelling strategy. We used both the raw data of the initial problem and then the normalized data from the DEA

Table 1. Results obtained with DEA model and PROMETHEE II (normalized data)

Rank	Driver ID	Cross index scores (DEA)	Driver ID	Net flow scores (PII)
1	50	1.000	50	71.958
2	21	0.9783	54	67.751
3	24	0.9648	42	66.671
4	42	0.9498	11	66.26
5	9	0.9451	12	65.09
6	54	0.9419	20	63.584
7	20	0.9303	1	62.131
8	19	0.9199	8	61.244
9	48	0.9134	21	60.743
10	8	0.9079	48	60.266
11	12	0.8984	9	59.663
12	30	0.8912	19	59.392
13	11	0.8889	30	58.981
14	5	0.8762	44	58.749
15	44	0.8726	24	58.274
16	41	0.8712	46	57.931
17	1	0.8610	35	57.669
18	46	0.8505	5	56.705
19	39	0.8476	38	55.521
20	36	0.8382	39	54.904
21	14	0.8366	14	54.364
22	35	0.8240	36	53.733
23	40	0.8191	41	53.657
24	27	0.8055	26	53.343
25	47	0.7977	40	52.5
26	26	0.7894	34	52.102
27	55	0.7804	51	51.105
28	16	0.7728	27	50.969
29	25	0.7710	13	50.72
30	34	0.7694	25	49.537
31	37	0.7661	53	49.498
32	45	0.7579	45	48.694
33	7	0.7500	47	47.972
34	38	0.7468	4	47.608
35	51	0.7441	16	47.472
36	29	0.7391	15	46.37
37	4	0.7382	55	46.211
38	32	0.7301	7	44.745
39	13	0.7294	37	44.655
40	53	0.7098	17	44.496
41	2	0.6960	28	42.558
42	28	0.6907	32	41.593
43	22	0.6828	2	39.665
44	3	0.6825	23	38.785
45	17	0.6710	29	37.324
46	49	0.6683	22	36.964
47	15	0.6626	31	36.072
48	6	0.6609	6	35.923
49	23	0.6577	3	35.884

50	31	0.6505	49	33.711
51	18	0.6452	33	33.672
52	33	0.6193	43	33.667
53	43	0.5937	18	33.243
54	10	0.5871	10	31.208
55	52	0.5806	52	26.492

model to define the preferences functions and the corresponding thresholds. As Chatterjee and Chakraborty considered in their work (2014), the PROMETHEE method is very robust against the normalization scheme and the results with raw data and normalized data are almost the same (a spearman correlation value of 0.9923). Therefore, in this work we just report on the normalized data (as for the case of DEA) when presenting the PROMETHEE ranking. This comparative analysis with PROMETHEE would allow us to enhance the analysis of the best and worst solutions highlighted with the previous model. The results in terms of a ranking of the 55 drivers based on the DEA method and the PROMETHEE method (both based on normalized data) are presented in Table 1.

Subsequently, we calculated the Spearman's correlation coefficient to compare the ranking results of the DEA and PROMETHEE model. The value of 0.937 indicates a high correlation between the two rankings. We also tested the Kendall's correlation coefficient, which is 0.789. As expected, this value is smaller than Spearman's correlation but again high enough. This high replicability of the results underlines the robustness of the best (and worst) drivers. When focusing on the drivers ranked at the best positions by the DEA model (cf. Table 1), we observe that they are all ranked in the top positions of the PROMETHEE II ranking. The strongest difference concerns the driver with ID 24 that is ranked at the 3rd position with DEA but only at the 15th position by the PROMETHEE II model. Concerning the worst solutions, the correlation is very high between the two rankings. Based on these observations, the best and worst drivers seem very robust. Consequently, we may consider the drivers that belong to these two categories as representative alternatives of the problem. By analyzing more precisely their performances, we could identify the profiles of these representative drivers. In the case of the worst drivers, it may allow us to define strategies and actions to apply in order to improve some of their inabilities.

5. Conclusions

In this study, we applied a multiple layer DEA based composite indicator model to assess the driving performance of 55 drivers aged 70 and older. This model allowed us to aggregate the values of 16 indicators within a particular category of a particular layer by using a weighted sum approach. Then, the cross index method was used to rank all the drivers with respect to their global performance. In order to quantify the robustness of the ranking, we modelled the problem with PROMETHEE II and compared the results. The calculation of the correlation coefficient underlined the robustness of the final solutions.

To conclude, this study has shown the value of using a DEA model as well as a MCDA method (PROMETHEE) for drivers' evaluation. This approach allowed us to rank all the drivers based on their performance and to assess the robustness of the best and worst candidates. In future work, we will improve this study by considering the combination of DEA and PROMETHEE as an analyzing tool to give older drivers more insight in the characteristics of their driving performance. Applying Plan GAIA (Geometrical Analysis for Interactive Aid) (Brans and Mareschal, 2002) can be a case in this analysis. It also would be valuable to incorporate an artificially created, ideal driver in the analysis, so that instead of a relative comparison, an evaluation of drivers in a more absolute manner would be possible.

References

- Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdas, M., 2010. PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 200(1), 198–215.
- Belton, V. and Vickers, S. P. 1993, Demystifying DEA - A Visual Interactive Approach Based on Multiple Criteria Analysis, *Journal of the Operational Research Society*, 44(9), 883-896.
- Brans J.P., 1982. L'ingénierie de la décision; Elaboration d'instruments d'aide à la décision. La méthode PROMETHEE, In Nadeau R. and Landry M., Presses de l'Université Laval Québec, Canada.

- Brans, J. P. and Mareschal, B., 2002. PROMETHEE-GAIA: Une méthodologie d'aide à la décision en présence de critères multiples, Editions de l'université de Bruxelles et Ellipses, Belgium.
- Carr, D. B., Schwartzberg, J. G., Manning, L., and Sempek, J., 2010. "Physician's guide to assessing and counseling older drivers." 2nd edition. Washington D.C.: National Highway Traffic Safety Administration.
- Cherchye, L. et al., 2008. Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59(2), 239–251.
- Clark, P.G., Blissmer, B.J., Greene, G.W., Lees, F.D., Riebe, D.A., Stamm, K.E., 2011. Maintaining exercise and healthful eating in older adults: The SENIOR project II: Study design and methodology. *Contemporary Clinical Trials*, 32(1), 129–139.
- Cooper, W.W., Seiford, L.M. & Tone, K., 2007. Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software: Second edition.
- Cuenaen, A., Jongen, E.M.M., Brijs, T., Brijs, K., Lutin, M., Van Vlierden, K., Wets, G., 2015. The relations between specific measures of simulated driving ability and functional ability: New insights for assessment and training programs of older drivers. *Transportation Research Part F*. <http://dx.doi.org/10.1016/j.trf.2015.05.007>
- Despotis, D.K., 2005. A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society*, 56(8), 969–980.
- Doyle, J. and Green, R. 1993, 'Data Envelopment Analysis and Multiple Criteria Decision Making', *OMEGA*, 21(6), 713-715.
- Edwards, J.D. et al., 2005. Reliability and validity of useful field of view test scores as administered by personal computer. *Journal of clinical and experimental neuropsychology*, 27(5), 529–543.
- Folstein, M.F., Folstein, S.E. & McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3), 189–198.
- Freudenberg, M., 2003. Composite indicators of country performance: a critical assessment. *OECD Science, Technology and Industry Working Papers*, 16, 35.
- Gardner, M.M., Buchner, D.M., Robertson, M.C., Campbell, A.J., 2001. Practical implementation of an exercise-based falls prevention programme. *Age and ageing*, 30(1), 77–83.
- Hakamies-Blomqvist, L. E. (1993). "Fatal accidents of older drivers." *Accident Analysis and Prevention*, 25, 19–27.
- Hayez, Q., De Smet, Y. and Bonney, J., 2012, 'D-Sight: a new decision making software to address multi-criteria problems', *International Journal of Decision Support Systems Technologies*, 4, (4), 1-23.
- 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.
- Horswill, M.S., Kemala, C. N., Wetton, M., Scialfa, C.T. and Pachana, N.A., 2010. "Improving older drivers' hazard perception ability." *Psychology and Aging*, 25, 464-469.
- Keeney, R.L., Raiffa, H. & Rajala, D.W., 1979. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(7).
- Langford, J. and Koppel, S., 2006. "Epidemiology of older driver crashes – Identifying older driver risk factors and exposure patterns." *Transportation Research Part F*, 9, 309-321.
- Liu, C.C., 2003. Simulating Weights Restrictions in Data Envelopment Analysis by the Subjective and Objective Integrated Approach. *Web Journal of Chinese Management Review*, 6(1), 68-78.
- Lundberg, C. et al., 2003. The assessment of fitness to drive after a stroke: The Nordic Stroke Driver Screening Assessment. *Scandinavian Journal of Psychology*, 44(1), 23–30.
- Mareschal, B. and Brans, J. P., 1988. 'Geometrical representations for MCDA', *European Journal of Operation Research*, 34(1), 69-77.
- Mathieson, P. et al., 2013. SaMERU - Safer mobility for elderly road users. , (September), 317.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E., 2005. Handbook on constructing composite indicators: Methodology and user guide. Organisation for Co-operation and Development.
- OECD, 2008. Handbook on Constructing Composite Indicators,
- Rosser, D.A., Laidlaw, D.A. & Murdoch, I.E., 2001. The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *The British journal of ophthalmology*, 85(4), 432–436.
- Roy, B., 1985. *Méthodologie Multicritère d'Aide à la Décision*, Economica, Paris, France
- Roy, B., 1991. The outranking approach and the foundations of electre methods. *Theory and Decision*, 31(1), 49–73.
- Saaty, T.L., 1980. The Analytic Hierarchy Process, Available at: <http://www.expertchoice.com/markets/index.html#AHP>.
- Sexton, T.R., Silkman, R.H. & Hogan, J.A., 1986. Data envelopment analysis: Critique and extensions. *New Directions for Program Evaluation*, 32, 73–105. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/ev.1441/abstract>.
- Shang, J. and Sueyoshi, T., 1995. 'A Unified Framework for the Selection of a Flexible Manufacturing System', *European Journal of Operational Research*, 85, 297-315.
- Shen, Y.J., 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.
- Shen, Y.J., Hermans, E., Brijs, T., Wets, G., 2012. Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model. *Social Indicators Research*. 114 (2), 739-756.
- Takamura, Y. and Tone, K., 2003. A comparative site evaluation study for relocating Japanese government agencies out of Tokyo, *Socio-Economic Planning Sciences*, 37, 85–102.
- Trik, L. M., Toxopeus, R. and Wilson, D., 2010. "The effect of visibility, traffic density and navigational challenge on speed compensation and

- driving performance in older adults.” *Accident Analysis and Prevention*. 42, 1661-1676.
- Wechsler, D., 1955. “Manual for the Wechsler Adult Intelligence Scale.” Oxford, England: Psychological Corporation
- Yan, X., Radwan, E. and Guo, D., 2007. “Effects of major-road vehicle speed and driver age and gender on left-turn gap acceptance.” *Accident Analysis and Prevention*, 39, 843-852.
- Yılmaz, B. and Yurdusev, M. A., 2011, ‘Use Of Data Envelopment Analysis As a Multi Criteria Decision Tool – A case of irrigation management’, *Journal of Mathematical and Computational Applications*, 16(3), 669-679.
- Zhang, J., Fraser, S., Clarke, K. and Mao, Y., 1998. “Age-specific patterns of factors related to fatal motor vehicle traffic crashes: focus on young and elderly drivers.” *Public Health*, 112, 289-295.