Aalto University

Mat-2.4177 Seminar on case studies in operations research

Analyzing the efficiency of Finnish health care units

Client: THL

Yrjänä Hynninen (Project Manager)
Petri Ollikainen
Joona Putkonen
Antti Tenhola
Jussi Vähä-Vahe

04.05.2011

Table of Contents

1. Introduction

A growing literature on data envelop analysis (DEA) has emerged since the seminal paper of Charnes et al. (1978), offering numerous methods for examining the efficiency of decision-making units (DMUs). More importantly, according to Hollingsworth et al. (1999), there has been increasing interest in measuring the productive performance of health care services since the mid-1980s. Salo & Punkka (2011) argue that DEA models in health care give insights into which DMUs are more efficient than others when health indicators are viewed as outputs and when inputs consist of health-care investments and possibly contextual factors as well. For example, Garcia et al. (1999) analyze the efficiency of primary health units and explore how sensitive the DEA results are to the selection of output variables. In addition, Linna et al. (2010) have compared the performance of hospital care in four Nordic countries.

However, there are some problems related to DEA approach. Non-parametric methods, such as DEA, give the highest available efficiency score one for many units already with relatively small amounts of output/input –variables, leading to results with low value of information. In addition, in case of low number of observations, the efficiency frontier might be based on outliers causing results to be sensitive. This is the case specifically in small countries, such as Finland, where the number of comparable health care organizations is typically small. These impose significant challenges when comparing sufficient approaches and methods to study the efficiency of health care units, and imply that especially parametric statistical methods might be problematic.

The aim of this study is to analyze and test methods for comparing efficiency of different health care units in Finland. The analysis is done specifically considering the problems discussed earlier. The main research question of this study can be stated as follows: What is the most appropriate method for comparing the efficiency of health care units in Finland? The performance of analyzed methods is tested using the real data of Finnish dental health care units. The data is provided by THL and it is used to demonstrate what kinds of results different methods produce.

This research report is organized as follows. Section 2 presents the basic DEA model and its most used variations and their advantages and disadvantages. Also the recent ratio-based efficiency analysis model (REA) (Salo and Punkka, 2011) is introduced. Section 3 focuses on sensitivity analysis, mainly on bootstrapping method. Section 4 discusses how DEA is applied to health care sector, both nationally and globally, and what are the characteristics of this sector that the models need to consider. Section 5 presents the example results of different models and analyzes their applicability. Section 6 concludes.

2. DEA models

2.1. CCR model

The Charnes-Cooper-Rhodes model (CCR) is one of the most basic DEA models and was initially proposed by Charnes et al. in 1978. The basic idea of the CCR model is that it calculates the efficiency ratio for the DMUs based on their inputs and outputs and by comparing that ratio with other DMUs the model defines the efficiency of the DMU.

Assume that a DMU consumes M types of inputs and produces N types of outputs. The *k*th DMU consumes $x_{mk} \geq 0$ units of the *m*th input and produces $y_{nk} \geq 0$ units of the *n*th output. The preference information about the relative values of inputs and outputs is represented by nonnegative weights v_i and u_j, respectively. The virtual input of the DMU is $\sum_m v_m x_{mk}$ and the virtual output $\sum_n u_n y_{nk}$. We assume these measures to be strictly positive for all feasible weights.

Based on virtual inputs and virtual outputs we define the efficiency ratio for each DMU. The efficiency ratio is

$$E_k(u, v) = \frac{\text{virtual output}}{\text{virtual input}} = \frac{\sum_n u_n y_{nk}}{\sum_m v_m x_{mk}}$$

Using linear programming we determine the optimal weights, which maximize the efficiency ratio for each DMU. The optimal weights usually vary from one DMU to another. If the DMU's efficiency ratio is the best of all DMUs with some weights, the particular DMU is efficient and will have an efficiency score of one (100%). The efficient DMUs define an efficient frontier which serves as a point of reference in the evaluation of efficiency.

If the DMU's efficiency ratio is not the best of all with any weights, the DMU is inefficient. The score of an inefficient DMU is usually less than one and it represents how close to the efficient frontier the DMU can optimally be. For example, if the score of a DMU is 0.8, it means that the particular DMU can produce only 80 % of the outputs that an efficient DMU can produce with the same inputs. To become efficient, the DMU needs to produce 25 % more outputs with the same inputs ($1/0.8 = 1.25$). The scores are always calculated with the most favorable weights for each DMU. (Cooper et al., 2007)

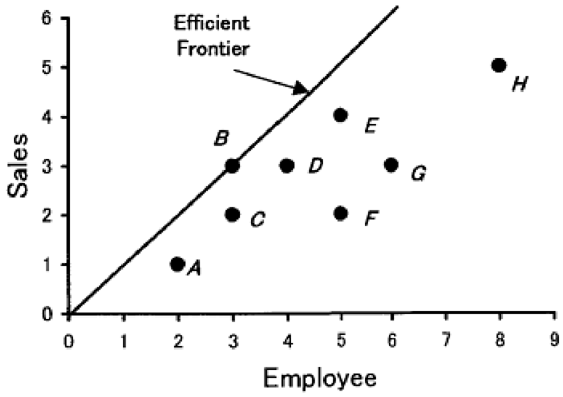


Figure 1: The CCR model

Figure 1 (Cooper et al., 2007) presents a simple example. The efficiencies of stores (A-H) are determined based on the sales and the number of employees. The store B is the only efficient DMU and determines the line of efficient frontier. The efficiency scores of other stores are calculated based on the relative distances from the frontier. For example, the efficiency score of A is 0.5 and the same of D is 0.75. The interpretation of the scores is that the sales of the store D are only 75 % of the sales of an efficient store. To become efficient, the store D needs to either increase sales by 33 % ($1/0.75 = 1.33$) or decrease the number of employees by 25 %.

We present here two different ways to calculate the efficiency scores. The first version aims to minimize inputs while satisfying at least the given output levels and the other version attempts to maximize outputs without requiring more of any input values. The versions are called input-oriented model and output-oriented model, respectively.

2.2. BCC model

The CCR model is built on the assumption of *constant returns to scale*, meaning that if all inputs are doubled, the output is also expected to double. The Banker-Charnes-Cooper (BCC) model, originally proposed by Banker et al. in 1984, is an extension of the CCR model. In their paper, Banker et al (1984) provide models for estimating both technical and scale efficiencies of DMUs. The BCC model takes into account that the productivity at the most productive scale size may not be attainable for other scale sizes at which a given DMU is operating. Therefore, the BCC model estimates the pure technical efficiency of a DMU at a given scale of operation.

The only difference between the CCR and BCC models is the convexity condition of the BCC model, which means that the frontiers of the BCC model have piecewise linear and concave characteristics, which lead to *variable returns to scale*. That is, the BCC model assumes *increasing, decreasing* and *constant returns to scale* at some point of the frontier (Figure 2). (Cooper et al., 2006, pp. 87-88)

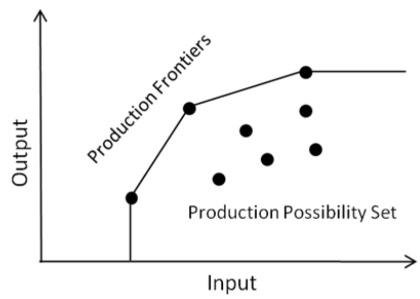


Figure 2: Production frontiers of the BCC model

Figure 3 (Cooper et al., 2007) illustrates the difference between the CCR model and the BCC model more precisely. The solid line, passing through A, B and C, represents the BCC model, whereas the dash line, passing through only B, represents the effective frontier of the CCR model. In general, the CCR efficiency cannot exceed the BCC efficiency. For example, let's calculate the BCC and CCR efficiencies of DMU D in Figure 3. The BCC efficiency is approximately $2.67/4 \approx 0.67$, whereas the CCR-efficiency of DMU D is smaller: $2.25/4 \approx 0.56$.

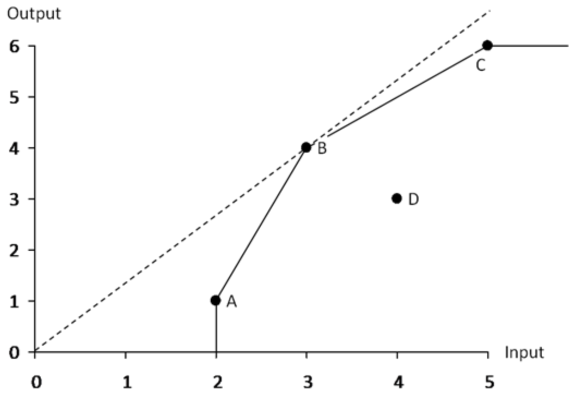


Figure 3: The BCC (solid line) model vs. the CCR (dash line) model

2.3. Weight restrictions

When using previous models (CCR, BCC), it is possible that there are many zeros in the optimal weights (v_i*,u_j*) of an inefficient DMU. A zero means that the particular input or output is effectively ignored in the efficiency evaluation and it is usually a sign that the DMU has a weakness in the corresponding input/output compared with other DMUs. The assumption of nonnegative weights is reasonable when we do not have any information about the relative values of inputs and outputs.

There are situations, however, where additional information is available and where one is willing to restrict the multiplier vectors v and u more than just by nonnegativity requirement. For these situations Cooper et al. (2007) present two methods: the assurance region method and the cone-ratio method. In the both methods the main idea is to restrict the feasible region of the weights.

The assurance region approach imposes constraints on the relative magnitude of the weights for special items. For example, we can add a constraint on the ratio of weights for input 1 and input 2 as follows:

$$L_{1,2} \leq \frac{v_2}{v_1} \leq U_{1,2},$$

where $L_{1,2}$ and $U_{1,2}$ are lower and upper bounds of the ratio v_2/v_1.

In the cone-ratio method we restrict the feasible region of weights to be in the convex cone generated by admissible nonnegative direction vectors. This method can be regarded as a generalization of the assurance region approach. It can deal with all the elements of input/output at the same time when in the assurance region method we have to define an equation for every pair of input/output elements.

Generally, by adding these constraints the efficiency score of a DMU is reduced and a DMU previously characterized as efficient may be found to be inefficient. Therefore, one has to be careful in choosing these bounds.

2.4. REA

When using the traditional DEA models, the efficiency scores of DMUs' represent the best possible efficiencies using the weights that are most favorable to each DMU. Traditional results do not include information how the efficiency score changes when using different input/output weights even though other weights may reflect relevant situations.

Salo and Punkka (2011) present a new model for analyzing efficiencies: the ratio-based efficiency analysis (REA). In REA model the efficiency scores of DMUs' are evaluated with every feasible weight combination. One can then examine for example how the efficiency scores change or what the ranking intervals of the DMUs are. With this information the decision maker sees if the efficiency of a DMU is robust.

Also the efficiency dominance can be examined. DMU_k dominates DMU_l if the efficiency ratio of DMU_k is at least as high as that of DMU_l for all feasible weights, and moreover, there exist some weights for which its efficiency is strictly higher. With this kind of examination the efficiency comparison can easily be made.

Another advantage of the REA model is that it is not as sensitive to outliers as the basic DEA models. In the basic models the introduction or removal of a single outlier may shift the efficient frontier drastically and thus disrupt efficiency scores. Using the REA ranking intervals, the similar manipulation of the sample changes the rankings at most only by one unit. The REA model is also suitable for situations where the number of DMUs is small, because the results are not computed relative to an efficient frontier.

3. Sensitivity analysis

3.1. Bootstrapping

3.1.1. Method

The measures of efficiency are relative ones and provide means for ranking different decision making units (DMUs), i.e. health care units in the current study. Therefore, it is also important to analyze the sensitivity of the estimated efficiencies to the sampling process. One method to address this issue is a bootstrap-method. The introduction of the bootstrap-method dates back to Efron (1979), when he showed that the method works satisfactorily on a variety of estimation problems, such as the estimation of the variance of the sample median and confidence intervals. As noted by Xue and Harker (1999); although the bootstrap is a computationally intense method, the modern computer is, however, more than sufficient for the computation required.

The problem solved by the bootstrap is mainly an estimation problem, and the principle of the method is simple and straightforward. Consider a random sample $X = (X_1, X_2, \ldots, X_n)$ from a population with an unknown distribution *F*. The goal is to estimate the sampling distribution of some pre-specified random variable *R(X,F)*, based on the real data set x, where $x = (x_1, x_2, \ldots, x_n)$ denotes the observed realization of $X = (X_1, X_2, \ldots, X_n)$. (Efron, 1979, Xue and Harker, 1999) To clarify this, bootstrap-method is a computer-based method for assigning measures of accuracy to sample estimates by constructing a number of resamples of the observed dataset with equal size, each of which is obtained by random sampling with replacement from the original dataset.

To illustrate this, consider the following coin-flipping example. We flip the coin and record whether it lands heads or tails. Let $X = x_1, x_2, \ldots, x_{10}$ be 10 observations from the experiment, where $x_i = 1$ if the coin lands heads, and 0 otherwise. Instead of using normal theory, such as t-statistic, we can use the bootstrap method and resampling to derive the distribution of sample mean $\overline{x}$. First, a bootstrap resample can be derived by resampling the data randomly with replacements. A first resample could look like for example like this: $X_1^* = (x_1, x_6, x_5, x_2, x_1, x_{10}, x_4, x_3, x_7, x_{10})$. It is important to notice that the resample can include duplicates, and the number of data points in the bootstrap resample have to be equal to the number of data points in the original observation data. The first bootstrap mean μ_1^* can be now easily calculated. By repeating this procedure for several times, an empirical bootstrap distribution of sample mean can be derived. This distribution can then be used for further analysis.

3.1.2. Applications

Since the introduction, the bootstrap has quickly become a popular and powerful statistical tool used to address some problems in statistical analysis. Simar (1992) was the first to use the bootstrap method in the non-parametric frontier analysis. He notes that in almost all cases, the sampling distributions are not available due to the non-linearity of the estimation procedures or the lack of parametric distributional assumptions on the residuals, and implies that in this case bootstrapping can help to get an insight on those issues. He concludes that the bootstrap is an appealing tool in the context of frontier estimation and efficiency analysis. The method provides not only a means to analyze the sensitivity of the ranking of the different units in terms of their inefficiency, with a measure of the statistical significance of the difference between the efficiencies, but it also provides proxy for the sampling distribution of estimators when analytical results are not yet obtained.

On the other hand, Atkinson and Wilson (1995) have used the bootstrap to construct the confidence intervals for the means of the DEA efficiency scores. Furthermore, they note that the bootstrap is one of several resampling methods which employ Monte Carlo techniques to approximate the small-sample distributions of estimators. In addition, they agree with Simar (1992) that the method is often of particular use in cases where analytic results are not available. Similarly, Ferrier and Hirschberg (1997) have derived the confidence intervals and a measure of bias for the DEA efficiency scores.

Simar and Wilson (1998) have utilized bootstrapping in analyzing the sensitivity of the DEA efficiency scores related to the variations of the estimated frontier. They clarify our discussion by stating that the bootstrapping is based on the idea of repeatedly simulating the data-generating process (DGP) usually through resampling, and applying the original estimator to each simulated sample. This way the resulting estimates mimic the sampling distribution of the original estimator. However, the key and also the greatest challenge, is to clearly define the DGP, i.e. the function that is supposed to characterize the entire population from which the data set has been drawn. Finally, Xue and Harker (1999) use the bootstrap-method to obtain a theoretically appropriate solution to the problem posed in the regression analysis of the DEA efficiency scores due to the inherent dependency among the DMUs' efficiency scores.

Grosskopf (1996) has reviewed literature regarding the bootstrapping, and finds the idea of using the bootstrapping to construct confidence intervals extremely appealing. She also considers that Simar and Wilson (1998) have suggested a very reasonable way of thinking about the DGP, and provided innovative solutions to the boundary and bias problems involved with applying bootstrapping to non-parametric technical efficiency.

Efron and Tibshirani (1986) have addressed the issue of how many bootstrap-replications one must take. They conclude that for the bootstrap-estimate of standard error already 25 Monte Carlo replications give reasonable results, whereas there is little improvement past 100 replications. On the other hand, they point out that confidence intervals are a fundamentally more ambitious measure of statistical accuracy than standard errors. Based on calculations of Efron (1987), Efron and Tibshirani (1986) suggest a rough minimum of 1000 for the number of Monte Carlo bootstraps necessary to achieve the bootstrap confidence intervals. Smaller values such as 250 replications can be somewhat useful for calculating percentile intervals.

More recent studies have also applied the bootstrapping in studying different health care units. For example, Staat (2006) presents results of a research using a DEA-bootstrap approach to study the efficiency of hospitals in Germany. He states that efficiency estimates based on DEA-type methods are biased upwards, and the bias depends on sample size N as well as on the curvature of the frontier and the magnitude of the density at the frontier. Furthermore, he proposes that in order to obtain bias corrected estimates for the multiple-input-multiple-output case, the bootstrap method must be applied. In addition, Medin et al. (2010) estimate cost efficiency scores for the performance of university hospitals in the Nordic countries by using the bootstrap bias-corrected procedure.

Based on the literature review, the bootstrap is considered to be specifically useful in cases where the sampling distributions and analytic results are not available, or when the sample size is small. The method is also considered to be simple and straightforward to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, and providing means to control and check the stability of the results. The recommendations of the situations when to use the bootstrap procedure suggested by Adèr et al. (2008) are consistent with the findings discussed above, and are summarized in the Table 1**.**

Table 1: Situations where bootstrapping procedures are useful

When the theoretical distribution of a statistic of interest is complicated or unknown.
Since the bootstrapping procedure is distribution-independent, it provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution.
When the sample size is insufficient for straightforward statistical inference.
If the underlying distribution is well-known, bootstrapping provides a way to account for the distortions caused by the specific sample that may not be fully representative of the population.
When power calculations have to be performed, and a small pilot sample is available.
Most power and sample size calculations are heavily dependent on the standard deviation of the statistic of interest. If the estimate used is incorrect, the required sample size will also be wrong. One method to get an impression of the variation of the statistic is to use a small pilot sample and perform bootstrapping on it to get impression of the variance.

On the other hand, the bootstrap-method does have also limitations. Campbell and Torgerson (1999) explain that many of the criticism presented in the literature are related to the simplicity of the assumptions of the model (Briggs et al., 1997, Campbell and Torgerson, 1997, Mooney and Duval, 1993). Therefore, the method may conceal the fact that many important assumptions are being made, such as independence of samples, when undertaking the bootstrap analysis.

Based on the literature review, it can be concluded that the bootstrap-method is applicable for sensitivity analysis in non-parametric frontier analysis, and for constructing confidence intervals for the DEA efficiency scores. The method has been applied also multiple times in studying different health care units. These findings suggest that the bootstrap-method is highly potential for studying the efficiency of health care units in Finland.

3.2. Other methods for sensitivity analysis

Bootstrapping is not the only method for sensitivity analysis. As Atkinson and Wilson (1995) imply, also other resampling methods exist that employ Monte Carlo techniques, such as jackknifing, cross-validation, and permutation tests. Completely other approach to sensitivity analysis is to consider the degrees of freedom in the envelopment model. As Cooper et al. (2006) notes, the number of degrees of freedom will increase with the number of DMUs and decrease with the number of inputs and outputs. On the other hand, algorithmic approaches relate to the use of algorithms that avoid the need for additional matrix inversions when generating solutions in the usual simplex algorithm computer codes (Cooper et al., 2006, Charnes et al., 1984, Charnes and Cooper, 1968). The basic idea of metric approaches is to use concepts such as distance or length in order to determine "radii of stability" within which the occurrence of data variations will not alter a DMU's classification from efficient to inefficient status (Cooper et al., 2006). Finally, multiplier model approaches are used in cases where the DMUs are numerous and it is not clear which ones require attention, unlike the other approaches above that treat one DMU at a time (Cooper et al., 2006, Thompson et al., 1994, Thompson et al., 1996). However, due to the scope of this study, these methods are not discussed here any further.

4. Applying DEA to health care

4.1. Literature review - previous applications

During the 1990's Data Envelopment Analysis rapidly became an acceptable method of efficiency analysis. Seiford (1994) lists 472 published articles and Ph.D. dissertations in his DEA bibliography already 1992. Less than a decade later Tavares (2003) includes almost 3200 items in his bibliography. Ray (2004) points out, however, that as instant as the success of DEA in management science was, in economics the welcome has been much less enthusiastic due to its shortcomings.

DEA models have been used in healthcare analysis since the early 1980's. Hollingsworth et al. (1999) reviewed non-parametric studies of health care efficiency made up to the end of 1997. Number of studies showed sharp increase towards the end of the period, half of them being published between years 1994 and 1997. Over 60% of the total 91 studies used DEA as the only method of analysis. Roughly a quarter used regression analysis in addition, mainly to regress factors on the efficiency scores in an attempt to identify the determinants of efficiency. Only one tenth of the studies combined DEA with newer developments such as Malmquist index or the use of efficiency scores as the dependent variable in secondary regression analysis. Studies of the early times are characterised of being limited with only basic DEA, since the methodology was still new and developing.

Later Hollingsworth (2003) updated the review with latest publications including studies up to 2002. Now the total number of studies was 188 and the trend was still growing. Figure 4 illustrates the growing popularity of DEA in healthcare efficiency analysis.

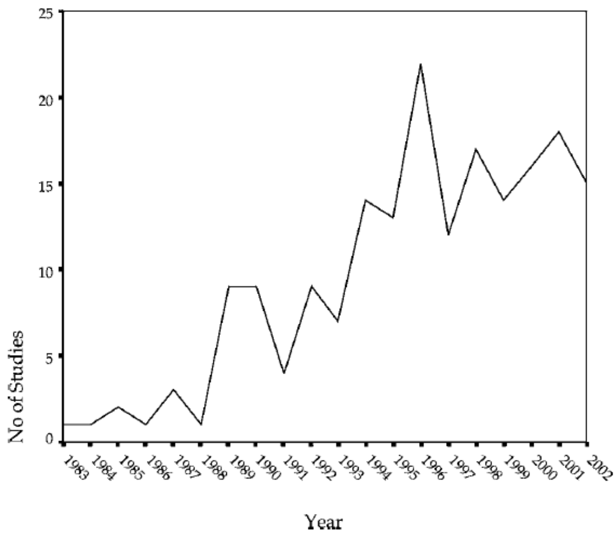


Figure 4: Number of DEA studies in health care efficiency analysis

With the latest research included the percentage of studies using DEA alone had dropped to 50. Proportion of the studies using DEA with regression remained in one quarter. Malmquist index and SFA, however, had gained significantly more popularity and represented now already one quarter of all studies. Hollingsworth notes that only a small part of the studies use weight restriction models or analysis of returns to scale. Similarly, only small part uses statistical or sensitivity analysis. These additions he sees as a major opportunity for improvement for future studies. Non-parametric techniques, such as stochastic frontier analysis, also gained popularity over the period of the two studies.

Overall finding of Hollingsworth was that in single sector studies DEA figures of public sector are higher than in private sector, indicating less variance in public sector. When compared across the whole healthcare sector, public hospitals still get higher scores, indicating in addition higher efficiency. Because of the large variance in research methods and on-going development of the techniques, Hollingsworth calls, however, for caution in comparisons.

As an example of a single study, Tsai and Molinero (2002) studied DMUs consisting of several activities and found that it is possible for a DMU to appear to be operating efficiently with constant returns to scale, but this may hide scale inefficiencies when the individual activities are studied separately. They further developed returns to scale analysis by incorporating a possibility of variable returns to scale for each one of the component activities a DMU is involved in. Two formulations were developed: The first considers the best overall allocation of resources between all the DMUs involved in the analysis and the second considers inefficient DMUs aiming to produce the best internal allocation of resources between the component activities in order to minimise overall inefficiency. Authors note that their model relies on priority judgments and thus includes an element of subjectivity. Returns to scale analysis appear, however, to be robust to the choice of weights in the model.

4.2. Returns to scale

The use of variable returns-to-scale assumption has to be very well argued. According to the point of view, which supports the use of assumption, there are situations where the scale of a unit has a specific influence on the unit's productivity (Cooper et al., 2007). Based on this point of view, the medium-sized units usually have the possibility to be more efficient than the very small or very large units. This means that the marginal productivity of the medium-sized units is relatively high; with one additional unit of input, they can produce more output than the other units. Especially the large units can be significantly less efficient than the smaller units. This idea is implemented from economics and it is reasonable to consider taking it into account when analyzing various-sized units. For example, with BCC model, one can compare the units from different scales at the same time and get useful results.

However, one has to be sure that the main reason for the poor performance of large units is the scale and that one cannot do anything to change that situation. The particular assumption accepts the weakness of large units as a fact and does not encourage finding a solution for that. There surely are some ways to develop the operations of large units to be better comparable with medium-sized units.

Also, there might be other reasons for the poor performance of large units than the scale. Those reasons and problems easily remain hidden if one does not consider the situation closely. For example, in health care the largest units are often responsible to receive the most difficult cases. Those cases require more resources and thus the efficiency, calculated by the quantity of operations, decreases. Now, if the patient structure is not linearly dependent on the size of the unit, the BCC model gives flawed results.

In economics, there is also another idea about the scale and marginal productivity, called the economies of scale. It simply means that the larger the company, the smaller the marginal costs. This leads directly to the greater efficiency. This phenomenon is usually seen in commodity production and thus it cannot be straightforwardly applied to health care, which is mostly a service business.

The point of presenting the idea of economies of scale is to show that it is not clear that large units automatically are less efficient than smaller units. Additionally, the assumption of variable returns to scale is so powerful that, if applied without a proper research, it might result more damage than advantage. Therefore, in many cases, it is safer to use the constant returns-to-scale assumption. If the scale differences are large, the units can be categorized by their size and compared to the units of same category with constant returns-to-scale assumption.

4.3. Weight restrictions

The use of weight restrictions seems very reasonable. Without restricting the weights, the relative values of outputs and inputs can be unrealistic. Often the optimal weights include zero, which means that the particular input or output type is not considered in the calculation of efficiency score at all. In health care sector, an example is that in the calculation of efficiency, only the number of nurses is considered, and the number of doctors is irrelevant. With weight restrictions, one can set upper and lower limits for the relative values of inputs and outputs.

Setting the restrictions has to be done carefully, because the limits are likely to be seen in the optimal weights. The restrictions should be based on statistical analysis or expert opinion. In both cases it is recommended to start with conservative weights and proceed step by step to more radical weights, following all the time how the efficiency scores change.

4.4. Geographical differences

O'Neill et al (2008) reviewed 79 DEA studies from 1984-2004 and found that there are fundamental differences in the methods used in DEA healthcare research between Europe and North-America. European studies use on average significantly more sophisticated methodology than their American counterparts. In defining efficiency 52% of the European studies used the more encompassing allocative method (requiring the relative price information of inputs and outputs) whereas only 12% of the American studies did the same. Thus, a vast majority of American studies settled with the simple technical efficiency analysis. European studies were also more comprehensive within the time frame, with 60 % conducting a multi-year study compared to the corresponding figure of 25 % in American studies. American researches used in addition far less DEA together with other methods such as stochastic frontier analysis (SFA) or the Malmquist index. European studies used less input variables (on average 3.8) but in turn more outputs (5.4) compared to American researches (4,8 and 4.7 respectively). The study sizes showed a massive difference between the two continents: European studies covered on average 74 DMUs whereas Americans had on average 440.

DEA and SFA methods have been found to produce similar efficiency estimates when applied to European hospitals (Jacobs, 2001) but divergent results in the US (Chirikos and Sear, 2000). According to O'Neill et al (2008) this indicates that allocative inefficiency is more of a problem in the US than in Europe. Such inefficiencies arise when hospitals compete by purchasing expensive equipment to attract physicians and patients. While this strategy might be efficient locally, it is not optimal nationwide because it leads to excess hospital capacity and partially empty surgical facilities for some trendy procedures such as transplants. All in all, healthcare is much more centralised and regulated in Europe, and Health Authorities influence resource allocation, reimbursements and hospital priorities.

In 1983 America changed the hospital reimbursement model from inpatient-based to diagnosis-related group (DRG)-based system. In DRG inpatient cases are classified into clinical groups based on expected resource use. When reimbursements were no longer based on total costs but individual cases, it changed radically the motivation of hospitals to get rid of the patients. Hence, DRG-adjusted discharges became a natural choice for output, while use of inpatient days dropped steadily to reach zero percent soon after the new millennium. The similar course of action started in Europe a decade later, but between 2001 and 2004 still a half of the European studies used inpatient days as one of the outputs.

4.5. Quality vs. quantity

Most of the DEA studies in healthcare have used quantitative outputs in the models, and there have been only few studies trying to implement quality measures in the outputs. Of course, one reason could be the lack of validated measures how to evaluate quality. For example, if mortality rates are used as quality outcomes, hospitals treating the sickest or most severely injured patients will become inefficient compared to their peers. (Nayar and Ozcan, 2008)

Newhouse (1970) argues that quantity and quality are two commodities to which the resources may be allocated. Since the resources are limited, there is a quantity-quality trade-off. It is usually thought that increasing quality may require more labor and capital, whereas efficiency improvements may lead to poorer performance in terms of quality. Laine et al. (2005) have studied the association between quality of care and technical efficiency in long-term care, and they state that defining and measuring is a multidimensional and complex problem.

DEA can be used to measure both dimensions of healthcare performance: technical efficiency and quality. Nayar and Ozcan (2008) studied whether the growing concern that hospitals are improving their efficiency at the expense of quality is valid. Nayar and Ozcan analyzed first the efficiency of acute care hospitals using measures of quality in DEA. Then the results were compared to a DEA model that uses measures of technical efficiency as inputs and outputs.

Nayar and Ozcan found that hospitals producing quantitative outputs efficiently were also producing quality outputs efficiently. Two thirds of the 53 hospitals analyzed were performing poorly in terms of both efficiency and quality. In addition, none of the hospitals was technically efficient but inefficient with respect to quality. Thus, Nayar and Ozcan's study does not support the efficiency-quality trade-off. In fact, it might be quite the opposite as some of the technically inefficient hospitals were performing well when it comes to quality.

4.6. Choosing variables

Choosing the input and output variables is an important phase of efficiency analysis. In this project, we do not focus closely on this phase, because it would require much more substance and understanding of the health care sector than we currently have. However, we present here the rule of thumb (Cooper et al., 2007) considering the number of variables:

$$n \geq \max\{m \times s, 3(m + s)\},$$

where n is the number of units, m the number of input variables and s the number of output variables. If this equation does not hold in the efficiency analysis, quite a few units might be seen as efficient. However, the number of efficient units can be decreased by using weight restrictions for example.

Cooper et al. (2007) also recommend starting the analysis with a small number of variables and increasing the number one by one, following constantly how the results change. It is better to have few important variables than too many less important even though the model might not be as accurate.

5. Results of data analysis

5.1. Dataset

In order to illustrate the differences between different DEA model modifications and assess the applicability of different DEA models to efficiency measurement in healthcare, we were provided with a dataset by THL. The dataset covered the year 2010, included 34 individual DMUs, and contained several input and output variables from the area of dental care in Finland. DMUs were basically different geographical healthcare areas, such as cities, that are autonomous in their resource allocation decisions. The dataset is presented and can be observed in the Appendix 1.

For this report, the DMU names were coded according to their relative sizes. The letter in a DMU name represents the size category (the largest DMUs in category A and the smallest in category C). The trailing number distinguishes the DMUs within a category.[1]

5.1.1. Variables

The dataset consisted of 4 input variables and 7 output variables. The variables are presented in Table 2.

Table 2: Variables of the provided dataset

Input variables	Description
Total costs	Total financial costs of the year 2010
Number of dentists	Average number of dentists during 2010
Number of dental hygienists	Average number of dental hygienists during 2010
Number of dental assistants	Average number of dental assistants during 2010
Output variables	Description
Number of treated patients in age group s1*	
Number of treated patients in age group s2*	
Number of treated patients in age group s3*	
Number of treated patients in age group s4*	
Number of treated patients in age group s5*	
Number of treated patients in age group s6*	
Weighted sum of operations completed	Sum of operations completed in 2010. Operations were weighted by cost factor to take into account varying expenses between operation types.
***) Age order: s1 < s2 < s3 < s4 < s5 < s6 (the oldest)**	

[1] Classification was based on the output variable *weighted sum of operations completed* (introduced later in the text). DMUs having aforementioned variable's value over 200 000 were classified in category A, DMUs having the value between 100 000 and 199 999 were classified in category B, and rest of the DMUs were classified in category C.

Input variables

The obtained input variable set can be divided into two mutually exclusive subsets. Another subset includes three variables (*number of dentists, number of dental hygienists*, and *number of dental assistants*) that represent the input effort of three different personnel groups counted in average man-years according to each personnel group. The other input variable subset consists of only the *total costs* variable including the salaries of the three aforementioned personnel groups but also other dental care related costs.

Output variables

Similarly to our input variables, we have two mutually exclusive output variable subsets. The first consists of six output variables counting the number of treated patients according to their age group. The second subset consists of the variable *weighted sum of operations completed* including all the treatments counted in the first subset. Instead of measuring the treated patients in the aforementioned age groups, the variable in the second output variable subset counts individual operations that were completed in 2010. In addition, the counted operations are weighted by a cost factor that takes into account varying costs between different dental care operations (e.g. root treatment vs. tooth extraction).

5.2. Analyses

In order to illustrate the differences between different DEA models, we ran several example results from the provided dataset. We wanted to sustain some degree of simplicity and thus decided not to conduct all possible DEA analyses with all possible variable combinations. Therefore, we selected *numbers of different dental care personnel groups* as our main input variables and *weighted sum of operations completed* as our main output variable. However, for the sake of curiosity, we ran an additional CCR-I analysis in which we used *total costs* instead of *personnel numbers* as an input variable to compare the obtained results with each other.

5.2.1. Variable selection

For analyses, we had to choose between overlapping variables in our dataset. As mentioned before, selected input variables were 1) *Number of dentists*, 2) *Number of dental hygienists*, and 3) *Number of dental assistants*. Our decision to choose these variables instead of *total costs* was based on two reasons. Firstly, single input variable would not have allowed us to illustrate the feature of DEA that allows variable weights to vary in the positive real axis. Secondly, *personnel numbers* are indifferent to different wage levels between geographical areas. If we want dental care to be offered everywhere in Finland, wage levels should be taken as given and thus they should not be included in analyses.

For the output variable selection we had two different options. The first subgroup contained *numbers of treated patients* according to their age group. As described earlier in this report, DEA allows variable weights to vary in the allowed space. We did not want to base our results on data that allows DEA models to consider different age groups of patients as more important than other age groups. If we wanted to use these *numbers of treated patients* as our output variable, these age groups should have been combined into single variable containing all the treated patients no matter what is their age group. However, in the end we decided to choose *weighted sum of operations completed* as our main output variable due to the fact that it in a sense contains the aforementioned figures and in addition takes into account varying costs between different dental care operation types.

5.2.2. CCR-I

Basic CCR-I (input oriented CCR) results are presented in Appendix 2. The most notable detail is that 5 DMUs out of 33 are considered as efficient. The all five are categorized as small (C) or medium-sized (B) DMUs. Otherwise ranking does not seem to obey any specific pattern. The worst-performing DMU, A6, got an efficiency score of 0.604, which means that in relation to the efficiency frontier it can produce only 60.4% of output with the same amount of input. Average efficiency score in these results was 0.85997 with standard deviation of 0.1095.

5.2.3. CCR-I - Largest DMUs

As the DEA should be applied to a group of to some degree similar DMUs, we wanted to test how it affects the efficiency scores of large DMUs (category A) and their relative rankings in case only large DMUs were included in the analysis. Figure 5 and Table 3 below present the efficiency scores of the CCR-I analysis conducted only for the group of large DMUs. Table 3 also lists the efficiency scores received in the original CCR-I that included all the DMUs.

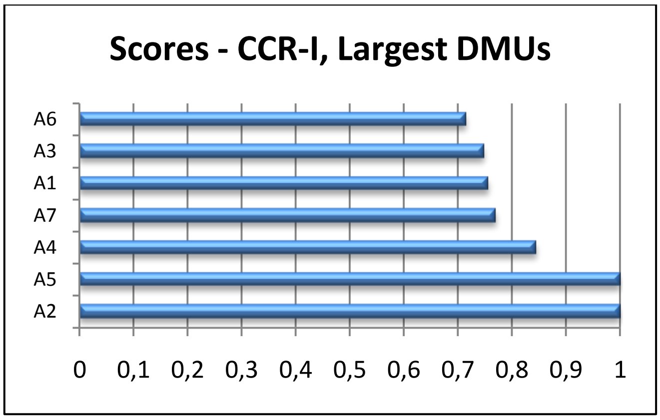


Figure 5: CCR-I efficiency scores of the largest DMUs

Table 3: CCR-I efficiency scores of the largest DMUs

DMU	SCORE	SCORE IN ORIGINAL CCR
A2	1	0,9734
A5	1	0,8809
A4	0,8448	0,7911
A7	0,7702	0,7256
A1	0,7562	0,7330
A3	0,7475	0,7170
A6	0,7153	0,6037

As we can see in the Table 3, large DMUs seem to receive consistently higher efficiency scores if the largest DMUs are compared only with each other. However, the ranking order within the group of large DMUs remains the same with only one exception: DMUs A7 and A1 change places. However, the efficiency difference between these two DMUs is relatively small in the both results.

5.2.4. CCR-I - Weighted sum of operations vs. total costs

Since the input variables in our dataset overlapped to some extent, we wanted to validate our results by producing similar CCR-I results where *total costs* would replace the *personnel* input variables. Intuitively it would be reasonable to expect somewhat similar results to original CCR-I results. These alternative CCR-I results are presented in Appendix 6.

If we compare these two sets of results, we can see that differences in efficiency scores and ranking vary quite a lot. The DMU C1 experiences the most radical change in its efficiency score and ranking. Its efficiency score decreases by 29.8 percentage points compared to the original CCR-I results. At the same time, its ranking decreases by 21 from 10th place to 31st. On average, the absolute value of the score difference is 11.2 percentage points, and the average absolute value of ranking difference is 7.5.

These results suggest that variable selection is one of the most essential parts of a DEA analysis.

5.2.5. BCC-I

Basic BCC-I results are presented in Appendix 3. Compared to our results from original CCR-I, BCC-I seem to double to amount of efficient DMUs. Now also the biggest DMUs seem to perform relatively better than in CCR-I model. Three DMUs from the category A received the best efficiency score, 1. Taking into account that the category A itself is quite small, as it contains only seven DMUs, large DMUs seem to perform relatively well when efficiency scores are determined using BCC model.

Another notable detail is that the worst performing DMUs perform a bit better than in the results determined with CCR-I. Average efficiency score in BCC-I results was 0.9039 which means that on average each DMU got 4.4 percentage points higher efficiency score compared to CCR results. Standard deviation of the scores was 0.1046.

The received BCC-I results are in line with the theory discussed earlier in this report. As BCC allows variable returns to scale, efficiency scores are at least as high as in CCR results.

5.2.6. Weight restricted BCC-I

In basic CCR and BCC models, variable weights are allowed to vary in non-negative real axis. Intuitively this allows efficiencies to be measured in points where importance differences (weight differences) between variables are irrelevantly large. In order to gain understanding on how deep impact variable weight restrictions have on results, we conducted another BCC-I efficiency measurement with the following input variable weight restrictions:

$$1 \leq \frac{Number\ of\ dentists}{Number\ of\ dental\ hygienists} \leq 5$$

$$1 \leq \frac{Number\ of\ dentists}{Number\ of\ dental\ assistants} \leq 5$$

$$0.5 \leq \frac{Number\ of\ dental\ hygienists}{Number\ of\ dental\ assistants} \leq 5$$

For example the first restriction means informally that in the measurement, dentists are considered at least as important as dental hygienists but at maximum five times as important. We do not state that these are the correct boundaries for the weights but instead we want to illustrate their effect on the results. The constraints were invented without expert opinions.

The results of Weight Restricted BCC are presented in Appendix 4. Compared to the basic BCC the number of efficient DMUs drops by three. On average DMUs are 4.15 percentage points less efficient as the average efficiency score drops to 0.8624 with standard deviation of 0.1122. Interesting detail is that every DMU in our sample receives the same or lower efficiency score than in the basic BCC results, even though it could have been possible to receive higher scores since irrelevant variable weights no more define the efficient frontier.

As a result of the weight restrictions, DMUs that performed significantly worse were the ones having a single personnel group of relatively small size. In the basic BCC these DMUs received higher cores without justifications due to the BCCs ability to emphasize that single input variable over the others.

Compared to basic CCR results, average efficiency scores are quite close to each other. Major difference is that Weight Restricted BCC allows also larger DMUs to be considered as efficient entities, even though the relative output/input ratio is smaller in comparison with smaller efficient DMUs.

5.2.7. REA - Ranking intervals

Also REA's ranking intervals feature was applied to the obtained dataset. Produced results can be observed in the Appendix 5. When generating REA ranking intervals results, the same variable weight restrictions were in place as in the WR-BCC results. The figure in the Appendix 5 demonstrates well how illustrative charts REA ranking intervals feature can produce to support decision making. The shorter the bar the more reliable the best ranking is in relation to other DMUs.

5.3. Discussion

In general our derived results are in line with the theory discussed earlier in the report. Due to the small sample size, no comprehensive implications can be made but important details can be emphasized.

In our results BCC generated higher efficiency scores than CCR, and in addition allowed large DMUs to produce smaller amount of output in relation to smaller DMUs. The decision between CCR and BCC is in our opinion political rather than technical - do we want to accept that large DMUs cannot achieve as high output/input ratios as small DMUs, and let that affect our decision making?

BCC in a sense compares DMUs to other DMUs which consume approximately similar amount of input. We applied the same principle to our sample in CCR model by comparing only large DMUs to each other. Even though large DMUs received higher efficiency scores, their relative rankings and efficiencies remained within the group of large DMUs quite much the same. The results did not give a strong signal that dividing a sample to subgroups of DMUs according to their sizes would add any value to the efficiency measurement.

Our view is that there is no good reason not to exploit the option of using weight restrictions. Excluding clearly irrelevant weight combinations will increase the quality of the results. Intuitively it seems that the higher the number of variables the more important the weight restrictions are to prevent too large emphasis on a single variable.

Our test of using different input variables representing partly the same input efforts resulted in very different efficiency scores and rankings. This indicates that variables are to be chosen with great consideration.

Finally, if the CCR model (constant returns to scale) is preferred over the BCC (variable returns to scale), then the REA model should be considered also. The REA is based on the assumption of constant returns to scale and produces the same results as the CCR but in addition provides very illustrative ranking intervals and pair dominances.

6. Conclusion

The purpose of this study was to help THL to identify the most appropriate DEA models for analyzing the efficiency of Finnish health care units. The aim was not only to compare the applicability of different DEA models but also to understand more deeply the characteristics of each model. In addition, the applicability of the bootstrap method for sensitivity analysis was studied, and the applicability of the REA model for producing more robust results was examined.

During this study, all the objectives presented in the beginning of the study have been accomplished. First, comprehensive literature review was conducted both on the mathematical background of DEA models and on efficiency studies in healthcare sector in general. This enabled us to identify potential DEA models, gain basic knowledge about the field, and to start working with the data provided by THL. The literature review focused specifically on DEA models of CCR, BCC (returns to scale) and weight restrictions, and on methods of bootstrapping and REA. Second, data analyses were conducted based on the data provided by THL. The data analyses enabled us to observe the applicability of different DEA models.

Based on the literature review and data-analyses conducted, it can be concluded that the answer to the research question is far from straightforward. First of all, it can be stated that the use of the variable returns-to-scale methods (for example BCC) can be justified if the aim is to compare different units of a very wide scale and it is sure that the scale of units affects the performance more than other external factors. However, the assumption of variable returns to scale is very strong and it can lead to flawed results if applied without a proper consideration. That is why constant returns-to-scale models (for example CCR and REA) are safer choices. If the scale is expected to be an important factor, one can divide the units to categories based on their size and use CCR within each category and get comparable results. However, if the CCR model is chosen for the efficiency analysis, it is better to switch completely to the REA model since it provides not only the same results as the CCR but also other information, such as ranking intervals, dominances, and efficiency intervals. Finally, the findings of this study support the use of weight restrictions if they are chosen carefully. By using conservative weight restrictions, irrelevant weight combinations can be eliminated.

According to the literature review, the bootstrap-method is applicable for sensitivity analysis in non-parametric frontier analysis, and for constructing confidence intervals for the DEA efficiency scores. The method has been applied also multiple times in studying different health care units. These findings suggest that the bootstrap-method is highly potential for studying the efficiency of health care units in Finland. However, the bootstrapping was not utilized in the data-analysis during this study, and no empirical evidence can be provided to support the preceding conclusions. Therefore, future research could study the applicability of the bootstrap-method even further by utilizing it in data-analysis of Finnish health care units.

References

ADÈR, H. J., MELLENBERGH, G. J. & HAND, D. J. 2008. *Advising on research methods: A consultant's companion,* Huizen:, Johannes van Kessel Publishing.

ATKINSON, S. E. & WILSON, P. W. 1995. COMPARING MEAN EFFICIENCY AND PRODUCTIVITY SCORES FROM SMALL SAMPLES - A BOOTSTRAP METHODOLOGY. *Journal of Productivity Analysis,* 6**,** 137-152.

BANKER, R. D., CHARNES, A. & COOPER, W. W. 1984. SOME MODELS FOR ESTIMATING TECHNICAL AND SCALE INEFFICIENCIES IN DATA ENVELOPMENT ANALYSIS. *Management Science,* 30**,** 1078-1092.

BRIGGS, A. H., WONDERLING, D. E. & MOONEY, C. Z. 1997. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics,* 6**,** 327-340.

CAMPBELL, M. K. & TORGERSON, D. J. 1997. Confidence intervals for cost-effectiveness ratios: the use of 'bootstrapping'. *Journal of Health Services Research & Policy,* 2**,** 253-5.

CAMPBELL, M. K. & TORGERSON, D. J. 1999. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. *Qjm-Monthly Journal of the Association of Physicians,* 92**,** 177-182.

CHARNES, A. & COOPER, W. W. 1968. STRUCTURAL SENSITIVITY ANALYSIS IN LINEAR PROGRAMMING AND AN EXACT PRODUCT FORM LEFT INVERSE. *Naval Research Logistics Quarterly,* 15**,** 517-&.

CHARNES, A., COOPER, W. W., LEWIN, A. Y., MOREY, R. C. & ROUSSEAU, J. 1984. Sensitivity and stability analysis in dea. *Annals of Operations Research,* 2**,** 139-156.

CHARNES, A., COOPER, W. W. & RHODES, E. 1978. MEASURING EFFICIENCY OF DECISION-MAKING UNITS. *European Journal of Operational Research,* 2**,** 429-444.

CHIRIKOS, T. N. & SEAR, A. M. 2000. Measuring hospital efficiency: A comparison of two approaches. *Health Services Research,* 34**,** 1389-1408.

COOPER, W. W., SEIFORD, L. M. & TONE, K. 2006. *Introduction to data envelopment analysis and its uses: with DEA-solver software and references,* New York:, Springer.

COOPER, W. W., SEIFORD, M. & TONE, K. 2007. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software,* New York:, Springer.

EFRON, B. 1979. 1977 RIETZ LECTURE - BOOTSTRAP METHODS - ANOTHER LOOK AT THE JACKKNIFE. *Annals of Statistics,* 7**,** 1-26.

EFRON, B. 1987. BETTER BOOTSTRAP CONFIDENCE-INTERVALS. *Journal of the American Statistical Association,* 82**,** 171-185.

EFRON, B. & TIBSHIRANI, R. 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science,* 1**,** 54-75.

FERRIER, G. D. & HIRSCHBERG, J. G. 1997. Bootstrapping confidence intervals for linear programming efficiency scores: With an illustration using Italian banking data. *Journal of Productivity Analysis,* 8**,** 19-33.

GARCÍA, F., MARCUELLO, C., SERRANO, D. & URBINA, O. 1999. Evaluation of Efficiency in Primary Health Care Centres: An Application of Data Envelopment Analysis. *Financial Accountability & Management,* 15**,** 67-83.

GROSSKOPF, S. 1996. Statistical inference and nonparametric efficiency: A selective survey. *Journal of Productivity Analysis,* 7**,** 161-176.

HOLLINGSWORTH, B. 2003. Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science,* 6**,** 203-218.

HOLLINGSWORTH, B., DAWSON, P. J. & MANIADAKIS, N. 1999. Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Management Science,* 2**,** 161-172.

JACOBS, R. 2001. Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. *Health Care Management Science,* 4**,** 103-115.

LAINE, J., FINNE-SOVERI, U. H., BJORKGREN, M., LINNA, M., NORO, A. & HAKKINEN, U. 2005. The association between quality of care and technical efficiency in long-term care. *International Journal for Quality in Health Care,* 17**,** 259-267.

LINNA, M., HAKKINEN, U., PELTOLA, M., MAGNUSSEN, J., ANTHUN, K. S., KITTELSEN, S., ROED, A., OLSEN, K., MEDIN, E. & REHNBERG, C. 2010. Measuring cost efficiency in the Nordic Hospitals-a cross-sectional comparison of public hospitals in 2002. *Health Care Management Science,* 13**,** 346-357.

MOONEY, C. Z. & DUVAL, R. D. 1993. *Bootstrapping: a nonparametric approach to statistical inference,* Newbury Park (CA):, Sage.

NAYAR, P. & OZCAN, Y. A. 2008. Data Envelopment Analysis comparison of hospital efficiency and quality. *Journal of Medical Systems,* 32**,** 193-199.

NEWHOUSE, J. P. 1970. TOWARD A THEORY OF NONPROFIT INSTITUTIONS - ECONOMIC MODEL OF A HOSPITAL. *American Economic Review,* 60**,** 64-74.

O'NEILL, L., RAUNER, M., HEIDENBERGER, K. & KRAUS, M. 2008. A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences,* 42**,** 158-189.

RAY, S. C. 2004. *Data envelopment analysis: theory and techniques for economics and operations research,* Cambridge, UK:, Cambridge University Press.

SALO, A. & PUNKKA, A. 2011. Ranking Intervals and Dominance Relations for Ratio-Based Efficiency Analysis. *Management Science,* 57**,** 200-214.

SEIFORD, L. M. 1994. A Data Envelopment Analysis Bibliography (1978-1992). *In:* CHARNES, A., COOPER, W. W., LEWIN, A. Y. & SEIFORD, L. M. (eds.) *Data envelopment analysis : theory, methodology and application.* Boston: Kluwer.

SIMAR, L. 1992. Estimating efficiencies from frontier models with panel data: A comparison of parametric, non-parametric and semi-parametric methods with bootstrapping. *Journal of Productivity Analysis,* 3**,** 171-203.

SIMAR, L. & WILSON, P. W. 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science,* 44**,** 49-61.

STAAT, M. 2006. Efficiency of hospitals in Germany: a DEA-bootstrap approach. *Applied Economics,* 38**,** 2255-2263.

TAVARES, G. 2003. A Bibliography of Data Envelopment Analysis (1978-2001). *Socio-Economic Planning Sciences*.

THOMPSON, R. G., DHARMAPALA, P. S. & DIAZ, J. 1994. Sensitivity analysis of efficiency measures with applications to Kansas Farming and Illinois coal mining. *In:* CHARNES, A., COOPER, W. W., LEWIN, A. Y. & SEIFORD, L. M. (eds.) *Data envelopment analysis : theory, methodology and application.* Boston :: Kluwer.

THOMPSON, R. G., DHARMAPALA, P. S., DIAZ, J., GONZALEZLIMA, M. D. & THRALL, R. M. 1996. DEA multiplier analytic center sensitivity with an illustrative application to independent oil companies. *Annals of Operations Research,* 66**,** 163-177.

TSAI, P. F. & MOLINERO, C. M. 2002. A variable returns to scale data envelopment analysis model for the joint determination of efficiencies with an example of the UK health service. *European Journal of Operational Research,* 141**,** 21-38.

XUE, M. & HARKER, P. T. 1999. *Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach,* [Philadelphia, Pa.].

Appendices

Appendix 1: Dataset

Appendix 2: Example results, CCR-I

Appendix 3: Example results, BCC-I

Appendix 4: Example results, BCC-I with weight restrictions

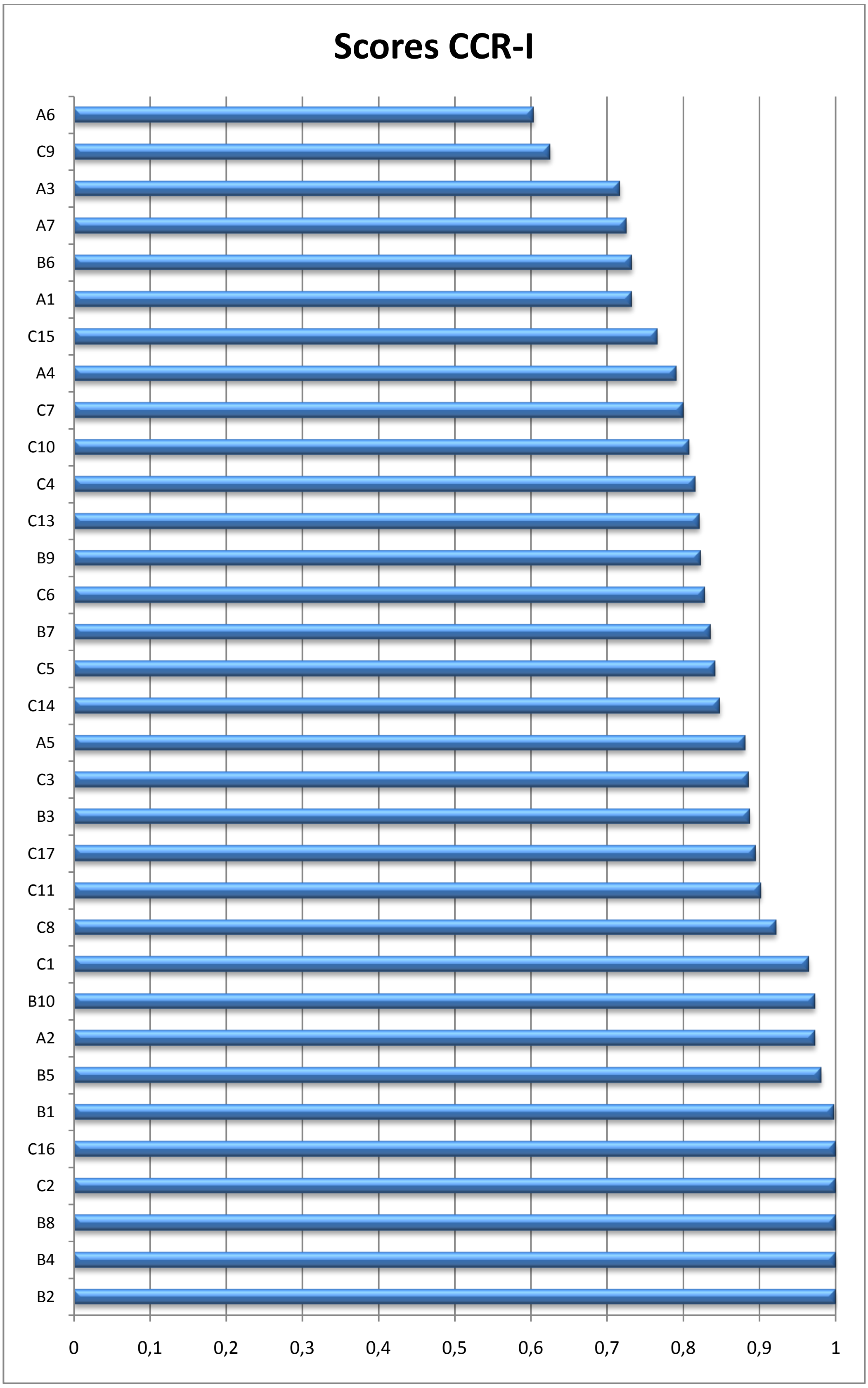
Appendix 5: Example results, REA Ranking Intervals with weight restrictions

Appendix 6: Example results, comparison of two sets of CCR-I results

Appendix 1: Dataset provided by THL

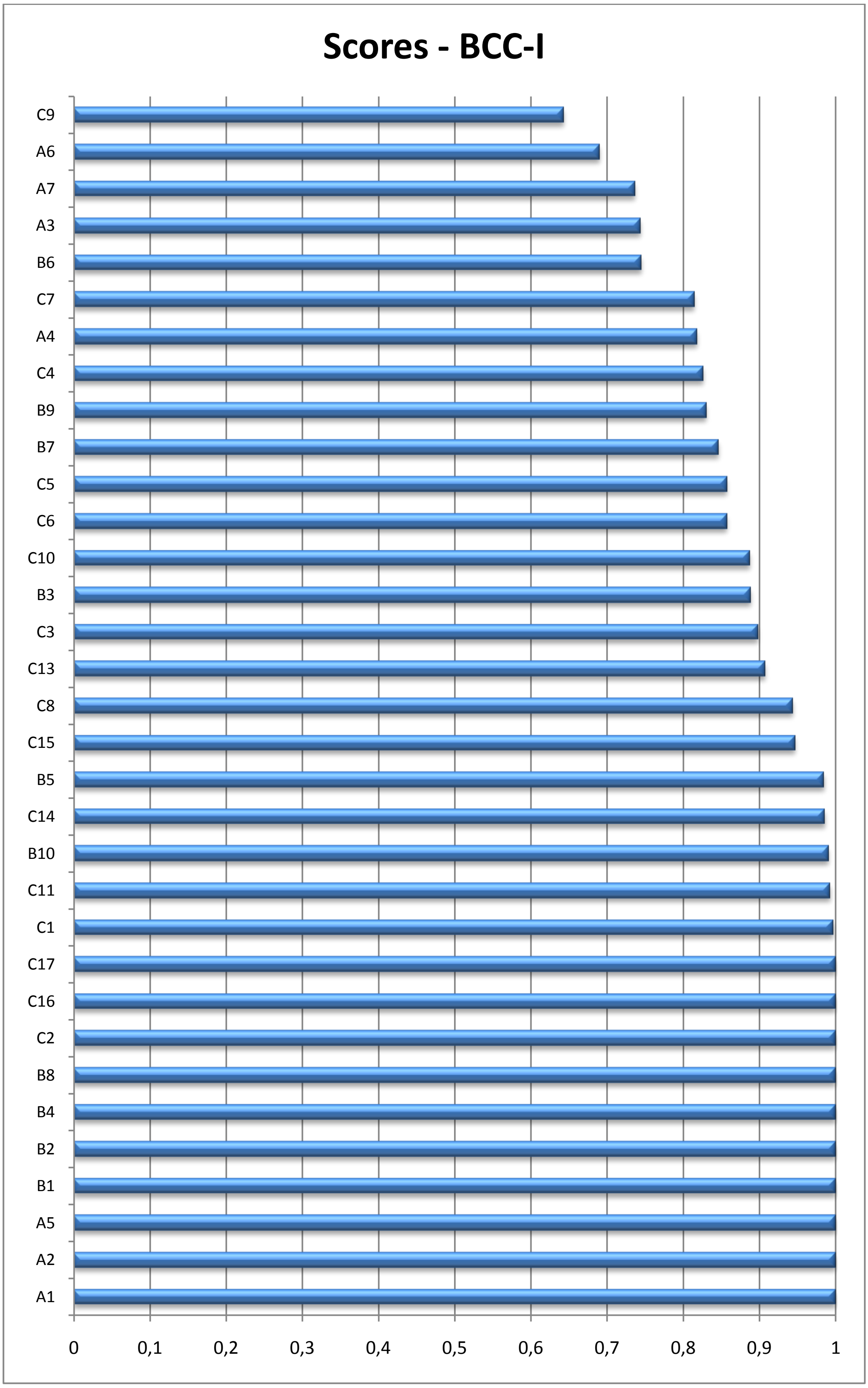
	OUTPUT							INPUT			
	Number of treated patients in age group si						Weighted sum of operations	Total costs	Dentists	Dental hygienists	Dental assistants
DMU coded	s1	s2	s3	s4	s5	s6					
A1	12842	42555	50984	23741	7302	6662	720093	44975290	153,0	72,2	325,3
A2	6668	20364	23226	11095	2988	1418	383983	17502985	61,0	29,3	97,8
A3	8940	22421	24223	10806	3564	2787	335847	19318817	74,3	33,5	125,4
A4	6450	16580	17018	7313	2044	2013	255905	12227155	52,8	22,1	83,1
A5	3716	13339	16712	5758	1767	1273	242502	10507410	51,5	15,0	82,0
A6	1	1	44689	1393	48	3	232321	10776823	69,0	22,0	99,0
A7	5107	14107	13581	6532	2075	1996	206113	9689196	46,4	19,4	104,2
B1	3213	9305	11602	6411	1682	1475	186201	6237909	28,2	16,4	45,0
B2	3140	9648	8411	5109	1760	1961	171424	5980705	25,4	13,5	42,4
B3	2314	7061	10662	5230	1464	1420	147874	6176656	24,8	14,0	41,0
B4	2283	7760	7053	4493	1018	223	124524	4016639	22,5	7,0	27,5
B5	2862	7832	7965	3257	1292	958	121078	5459737	20,9	7,9	29,8
B6	2634	6411	5201	3585	1488	988	114151	5224168	23,5	12,0	38,0
B7	2034	7261	6588	5957	2095	1638	105407	4611923	19,6	9,3	34,7
B8	1792	4946	4301	4152	1405	885	101510	4119416	17,0	7,0	21,0
B9	1316	5258	4401	2974	821	472	101301	4030731	20,9	8,0	27,3
B10	1659	5423	5857	4100	1149	819	100976	4067604	20,1	5,7	29,9
C1	1614	5202	5541	3868	1301	844	91062	4404700	20,3	5,3	23,3
C2	2286	5580	5170	3489	962	1053	90271	4003397	11,1	9,6	28,0
C3	1260	4063	4605	3475	997	656	85070	4058438	16,5	6,0	30,5
C4	2006	4927	3892	2918	990	725	84606	0	18,5	6,0	32,5
C5	1478	5418	4887	3373	1056	806	80949	3109573	16,8	5,8	26,4
C6	1711	4603	4678	2523	786	635	80322	3655996	16,0	6,6	22,3
C7	2096	5906	5143	1975	723	375	78929	3303651	17,2	6,0	28,3
C8	1973	4522	3470	1790	557	438	74428	2468876	12,3	8,6	19,2
C9	1757	4533	4421	2569	575	355	71949	3415670	20,0	7,0	31,0
C10	973	3551	4274	2443	721	340	66338	2645655	14,2	7,0	17,0
C11	1853	4711	3885	2516	745	458	65953	2747867	13,8	5,8	15,1
C12	1258	3962	3273	2309	711	554	63453	2668673	0,0	0,0	0,0
C13	1418	4429	3223	2431	618	597	60085	2598221	15,0	4,0	21,0
C14	878	2328	1987	1565	506	403	49912	1957640	9,7	5,8	12,5
C15	1129	2963	2810	1517	593	429	48061	1897774	10,5	5,0	13,0
C16	1019	2979	2788	1898	501	358	39455	1393346	8,0	2,0	18,0
C17	823	2070	1761	1582	524	453	33247	1448845	5,4	4,2	9,5

Appendix 2: Example results, CCR-I
Input variables: 1) number of dentists, 2) number of dental hygienists, 3) number of dental assistants
Output variables: 1) weighted sum of completed operations (operations weighted by cost factor)
Weight restrictions: none



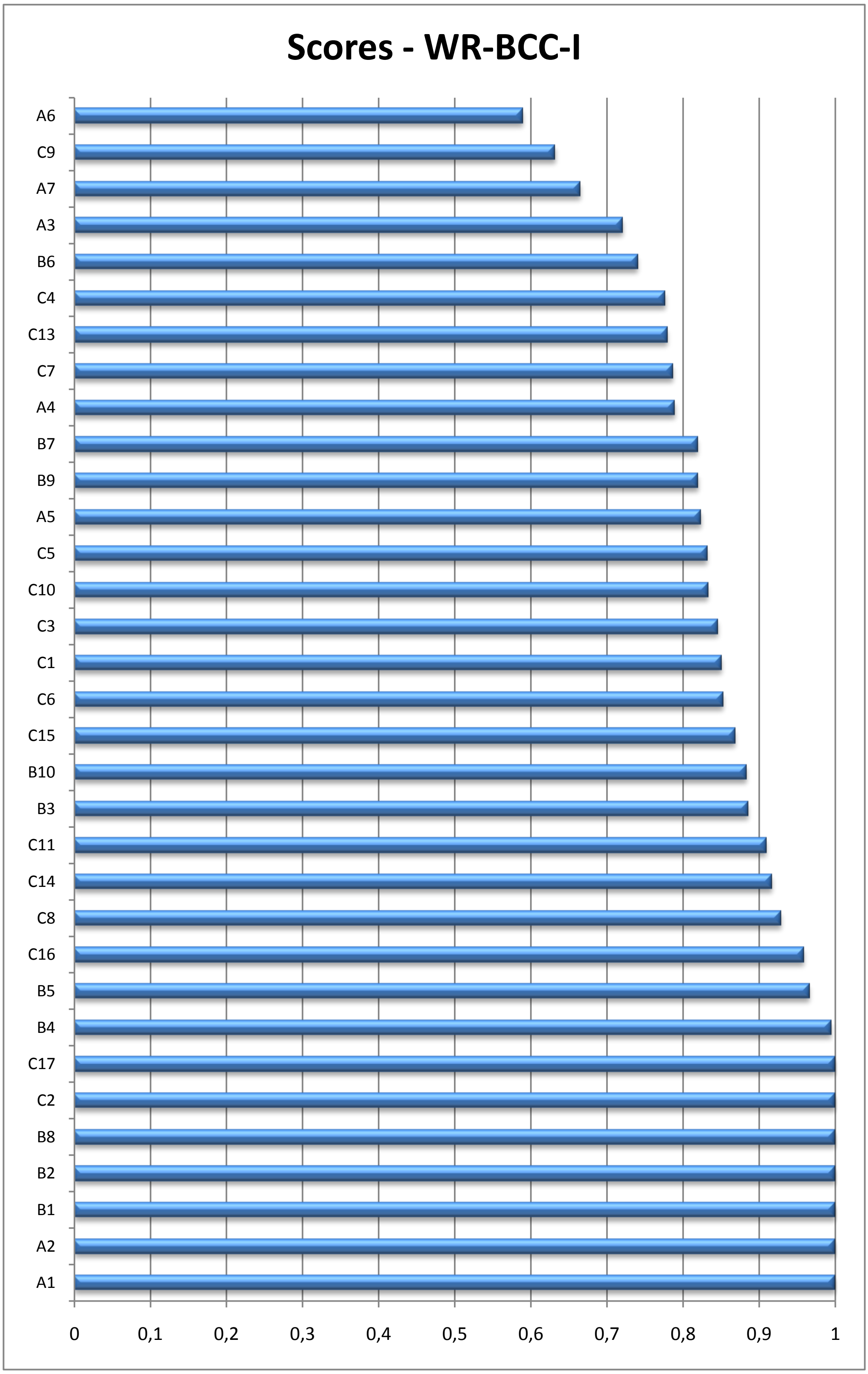
DMU	Score
B2	1
B4	1
B8	1
C2	1
C16	1
B1	0,9984
B5	0,9814
A2	0,9734
B10	0,9729
C1	0,9647
C8	0,9221
C11	0,9018
C17	0,8951
B3	0,8873
C3	0,8853
A5	0,8809
C14	0,8481
C5	0,8420
B7	0,8358
C6	0,8282
B9	0,8228
C13	0,8211
C4	0,8156
C10	0,8073
C7	0,8001
A4	0,7911
C15	0,7659
A1	0,7330
B6	0,7329
A7	0,7256
A3	0,7170
C9	0,6258
A6	0,6037

Appendix 3: Example results, BCC-I
Input variables: 1) number of dentists, 2) number of dental hygienists, 3) number of dental assistants
Output variables: 1) weighted sum of completed operations (operations weighted by cost factor)
Weight restrictions: none



DMU	Score
A1	1
A2	1
A5	1
B1	1
B2	1
B4	1
B8	1
C2	1
C16	1
C17	1
C1	0,9965
C11	0,9927
B10	0,9905
C14	0,9858
B5	0,9849
C15	0,9468
C8	0,9439
C13	0,9074
C3	0,8978
B3	0,8884
C10	0,8874
C6	0,8571
C5	0,8570
B7	0,8464
B9	0,8300
C4	0,8261
A4	0,8179
C7	0,8145
B6	0,7451
A3	0,7435
A7	0,7366
A6	0,6902
C9	0,6426

Appendix 4: Example results, BCC-I with weight restrictions
Input variables: 1) number of dentists, 2) number of dental hygienists, 3) number of dental assistants
Output variables: 1) weighted sum of completed operations (operations weighted by cost factor)
Weight restrictions:
1) 1 ≤ (dentists/assistants) ≤ 5; 2) 1 ≤ (dentists/hygienists) ≤ 5; 3) 0,5 ≤ (hygienists/assistants) ≤ 5



DMU	Score
A1	1
A2	1
B1	1
B2	1
B8	1
C2	1
C17	1
B4	0,9946
B5	0,9657
C16	0,9582
C8	0,9287
C14	0,9168
C11	0,9090
B3	0,8851
B10	0,8830
C15	0,8682
C6	0,8523
C1	0,8506
C3	0,8455
C10	0,8323
C5	0,8312
A5	0,8233
B9	0,8193
B7	0,8190
A4	0,7890
C7	0,7865
C13	0,7791
C4	0,7764
B6	0,7401
A3	0,7203
A7	0,6646
C9	0,6318
A6	0,5892

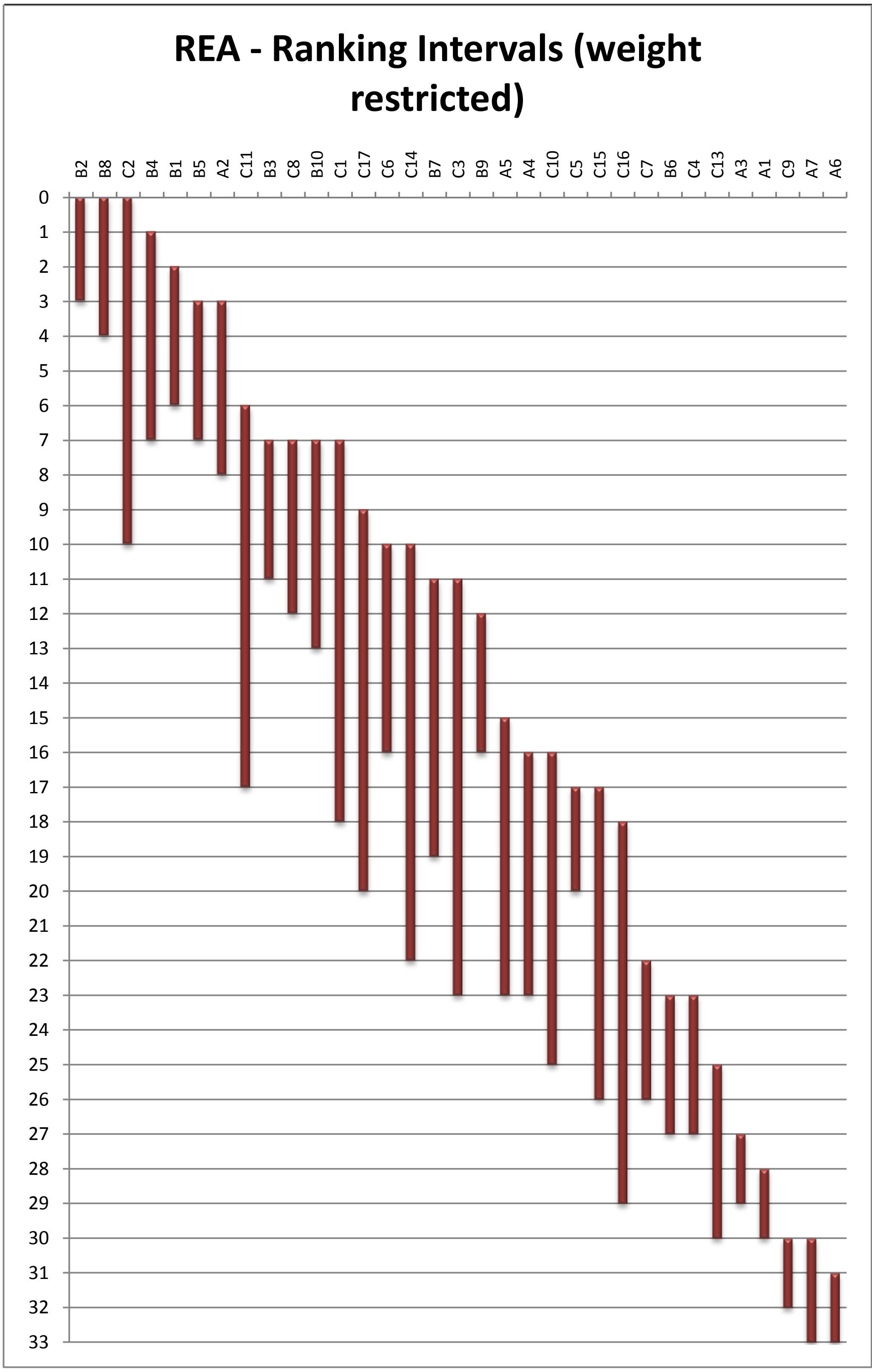
Appendix 5: Example results, REA Ranking Intervals with weight restrictions

Input variables: 1) number of dentists, 2) number of dental hygienists, 3) number of dental assistants

Output variables: 1) weighted sum of completed operations (operations weighted by cost factor)

Weight restrictions:

1) 1 ≤ (dentists/assistants) ≤ 5; 2) 1) 1 ≤ (dentists/hygienists) ≤ 5; 3) 1) 0,5 ≤ (hygienists/assistants) ≤ 5



DMU	Ranking	
	MIN	MAX
B2	1	3
B8	1	4
C2	1	10
B4	2	7
B1	3	6
B5	4	7
A2	4	8
C11	7	17
B3	8	11
C8	8	12
B10	8	13
C1	8	18
C17	10	20
C6	11	16
C14	11	22
B7	12	19
C3	12	23
B9	13	16
A5	16	23
A4	17	23
C10	17	25
C5	18	20
C15	18	26
C16	19	29
C7	23	26
B6	24	27
C4	24	27
C13	26	30
A3	28	29
A1	29	30
C9	31	32
A7	31	33
A6	32	33

Appendix 6: Example results, comparison of two sets of CCR-I results

Weighted sum of operations vs. total costs:

Input variables: Total costs; Output variables: Weighted sum of completed operations

Weighted sum of operations vs. personnel:

Input variables: Number of dentists, Number of dental hygienists, Number of dental assistants; Output variables: Weighted sum of completed operations

	Weighted sum of operations vs. total costs		Weighted sum of operations vs. personnel		Change in score	Change in ranking
DMU	Score	Ranking	Score	Ranking		
A1	0,5164	33	0,7330	28	-0,2166	+5
A2	0,7076	24	0,9734	8	-0,2657	+16
A3	0,5608	32	0,7170	31	-0,1562	+1
A4	0,6751	30	0,7911	26	-0,1160	+4
A5	0,7444	18	0,8809	16	-0,1365	+2
A6	0,6954	26	0,6037	33	+0,0916	-7
A7	0,6862	27	0,7256	30	-0,0395	-3
B1	0,9628	3	0,9984	6	-0,0356	-3
B2	0,9246	4	1	1	-0,0754	+3
B3	0,7722	14	0,8873	14	-0,1150	0
B4	1	1	1	1	0	0
B5	0,7153	22	0,9814	7	-0,2661	+15
B6	0,7048	25	0,7329	29	-0,0281	-4
B7	0,7372	20	0,8358	19	-0,0985	+1
B8	0,7948	12	1	1	-0,2052	+11
B9	0,8107	9	0,8228	21	-0,0121	-12
B10	0,8007	11	0,9729	9	-0,1721	+2
C1	0,6669	31	0,9647	10	-0,2978	+21
C2	0,7273	21	1	1	-0,2727	+20
C3	0,6761	29	0,8853	15	-0,2092	+14
C4	#N/A	#N/A	0,8156	23	#N/A	#N/A
C5	0,8397	6	0,8420	18	-0,0023	-12
C6	0,7087	23	0,8282	20	-0,1196	+3
C7	0,7706	15	0,8001	25	-0,0294	-10
C8	0,9724	2	0,9221	11	+0,0503	-9
C9	0,6795	28	0,6258	32	+0,0537	-4
C10	0,8088	10	0,8073	24	+0,0015	-14
C11	0,7742	13	0,9018	12	-0,1276	+1
C12	0,7669	16	#N/A	#N/A	#N/A	#N/A
C13	0,7459	17	0,8211	22	-0,0752	-5
C14	0,8224	7	0,8481	17	-0,0258	-10
C15	0,8169	8	0,7659	27	+0,0510	-19
C16	0,9134	5	1	1	-0,0866	+4
C17	0,7402	19	0,8951	13	-0,1549	+6