Analyzing the Efficiency of Dialysis Units through the use of Data Envelopment Analysis

Stephen Chu and Sara Dolcetti

A thesis submitted in partial fulfillment
of the requirements for the degree of

BACHELOR OF APPLIED SCIENCE

Supervisor: Dionne Aleman

Abstract

Data Envelopment Analysis (DEA) is a mathematical technique that can be applied to the methodology for multiple criteria decision analysis. It is used to measure the efficiency of multiple Decision Making Units (DMUs) which are calculated from a given set of inputs and their associated outputs. This paper discusses the application of DEA to dialysis centers in Canada through a comprehensive analysis of the results obtained from a DEA study of the Dialysis Registry in Toronto, Canada. With the small number of participating Toronto dialysis centers, results were inconclusive, as all centers were found to be efficient. Through further analysis, it was shown that with an increased sample size, DEA can be an effective measure of dialysis center efficiency. This paper is offered as a preliminary study, paving the way for future efficiency studies the benchmarking of Canadian Dialysis Centres using DEA.

Acknowledgements

We greatly appreciate the help offered by Mark Benaroia, the Staff Nephrologist at Grand River Hospital in Kitchener Ontario and the Toronto Dialysis Registry who provided the data for this study.

Contents

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Canada's publicly funded health care system is experiencing growing challenges due to rising demand and cost structures. National healthcare expenditures are currently approximately $171.9 billion, having increased 6.4%, or $10.3 billion, since 2007. Comprising 10.7% of the gross domestic product (GDP), this figure represents the largest ever recorded share of Canada's overall economic spending and is forecasted to continue rising over the coming years, primarily due to changes in demographics (i.e. population growth, aging) and services (i.e. advanced medical procedures) [Provincial and Territorial Ministers of Health, 2000] . Therefore, with operating expenditures that are growing faster than the economy [Canadian Institute for Health Information, 2005] , Canadian health services are in need of a mitigating strategy that reduces costs while maintaining a high level of patient care.

To address these two seemingly conflicting objectives, government efforts may be directed towards reducing the inefficiencies within the healthcare system so that the higher demand levels may be met with less financial aid. Consequently, analyses need to be performed to identify the sources of the inefficiencies before measures for improvement can be made.

While notably high hospital operational costs have prompted these centers to be the focus of efficiency studies, other services also significantly contribute to overall federal and provincial healthcare expenditures. Dialysis Treatment Centre services are one such service. Currently an estimated 8.3% of overall healthcare procedures, these figures are projected to grow at an increasing rate over the next several years due to the rising cases of diabetes and other renal ailments among Canadians.

A therefore important segment of the industry, this report focuses on analysing and comparing the efficiencies amongst Toronto dialysis centers and demonstrates the potential value of a mathematical technique called data envelopment analysis (DEA), in the process of identifying the sources of inefficiencies within the dialysis treatment system in order to meet the overall objective of improving the quality of dialysis delivery and care.

2 Dialysis Treatment in Canada

Canada operates under a universal health care system, which ensures that dialysis services are similar across treatment centers. One might expect uniform adherence to practice guidelines and uniform delivery of care both across the country and within provinces in a universal health care system. Yet, despite this universality, there is variability in the actual delivery of care that may be the result of differing physician attitudes and beliefs, resource allocations and individual patient states of health. This high level of variability may be due in part to the lack of a formal benchmark against which centers may evaluate themselves. To reduce variability, standardize care and ultimately improve center efficiencies, there is a need amongst Canadian Dialysis Centers for a benchmark to identify centers that are performing sub-optimally. Inefficient centers may then compare themselves against the benchmark to assist them in achieving similar levels of efficiency. DEA is an effective technique that can be employed to analyze the efficiencies of dialysis centers and establish such a benchmark.

3 Efficiency Measurement

Simply defined, efficiency is a ratio measurement of outputs to inputs and describes how well a unit is able to use its resources to produce a certain result.

$$Efficiency\,of\,Unit\,A = \frac{Output\,from\,Unit\,A}{Input\,to\,Unit\,A}$$

Efficiency studies may be performed using either a deterministic or stochastic approach, the difference between the two approaches being notable. Deterministic models assume that a unit's deviation from the ideal state (as determined in the model) is entirely attributed to inefficiency. Stochastic models attribute both inefficiency and random error to the observed deviation. Under the deterministic approach, the following techniques have traditionally been used: Ratio Analysis and Deterministic Frontier Analysis (an econometric method).

3.1 Ratio Analysis

Ratio analysis provides a simple, straightforward comparison between a single input and a single output. While the analysis is quick and completely empirical, its main disadvantage lies in the fact that it is limited to comparing one input to one output. Ratio analysis is also unable to assign relative weights based on importance of data. Examples of ratio analysis include quick and current ratios used in financial analysis.

3.2 Deterministic Frontier Analysis

Deterministic Frontier Analysis (DFA) is an econometric technique used in data analysis. Econometrics is concerned with empirically studying and testing the validity of economic theories using statistical analysis. DFA typically measures the efficiency of a single output as a function of several given inputs and parameters. Any deviation from the function in the actual output value is attributed to inefficiency.

The primary disadvantage of using DFA is that since it is an econometric technique, a distribution, such as linear or logarithmic, must be specified as the function of the model. If not enough information is provided, the distribution may be inaccurately chosen, leading to severe restrictions in analysis.

While Ratio Analysis and Deterministic Frontier Analysis both provide a good basis for studying the efficiency of a system, they do not rigorously resolve the issues of handling multiple inputs and outputs, separating input efficiency from variations between inputs and providing a clear means to distinguish between inefficient and efficient units [Nyman et al., 1989]. A technique that does incorporate these considerations, however, is Data Envelopment Analysis (DEA).

3.3 Data Envelopment Analysis

Introduced in 1957 by M. Farrell and later developed by Charnes, Cooper and Rhodes in 1978, DEA is a mathematical, non-parametric linear programming technique that measures efficiency based on multiple outputs and inputs. Consequently, it can be applied to the methodology for multiple criteria decision analysis [Charnes et al., 1978]. Commonly used in operations research and economics, DEA has been used in a number of studies to measure the technical efficiency of hospitals and other healthcare institutions (i.e. nursing homes) and has been quite successful in doing so [Kooreman, 1994]. A few studies have applied it to dialysis centers in Europe [Gerard et al., 2003][Kontodimopoulos et al., 2005], but none have investigated applying it to dialysis centers in North America. Therefore, this study aims to apply DEA to Canadian Dialysis Centers and to observe useful insights into their relative performances.

In practice, DEA is used to measure the efficiencies of multiple Decision Making Units (DMUs) which are calculated from a given set of weighted inputs and outputs. The most efficient DMUs are then identified and serve as an efficiency benchmark (formally referred to as the efficient frontier) for the entire set of DMUs.

Displayed graphically in Figure 1 (Based on Kooreman's DEA explanation in the paper 'Nursing Home Care in The Netherlands'), the points A,B,C,D and E each represent a different DMU with their inputs oriented along the x-axis and their outputs are oriented along the y-axis. The input and output variables along their respective axes are referred to as virtual inputs and outputs. This is due to the fact that they are both weighted sums of multiple values and do not represent actual individual figures.

In this figure, A, B and C are the most efficient units as they produce a given number of outputs with the fewest number of inputs (shown graphically as they are located on the outer edges of the set of all units). For example, units B and F both use the same amount of inputs (i.e., X_{BF}) but B produces a greater output than (i.e., y_B) than F does (i.e., y_F) and is therefore the most efficient unit for that particular value of x.

The line drawn through points O,A,B and C forms what is referred to as the efficient frontier or the isoquant and marks the optimal output quantities that are produced for differing input quantities. All units (termed frontier units) on the isoquant have an efficiency score of 1. Any unit that does not lie on the isoquant (termed a non-frontier unit) is regarded as inefficient and has an efficiency (in the case of unit F) of $\frac{X_{F1}}{X_{BF}}$. That is, for a given output, the ratio of the output's minimal x value to the x value associated with a given unit (where the minimal x value is the x value for which the line drawn perpendicular to the x-axis intersects the isoquant at the same output value as for the unit in question). The closer a unit lies to the isoquant, the nearer their efficiency score is to 1.

The example above describes the efficiency in an input oriented manner. This means that a non-frontier unit is characterised in terms of its failure to use minimal inputs to generate a given output. The opposite of this is output oriented efficiency in which a non-frontier unit is characterised in terms of its failure to generate a maximal output from a given input. For the purposes of this study, efficiencies were approached from an input oriented perspective as dialysis centers are better able to control their inputs (i.e., number of doctors, number of dialysis machines etc.) than their outputs (i.e., number of patients , quality of care etc.).

There are also multiple ways in which to calculate the efficiency scores of a non-frontier unit. Common alternatives include constant returns to scale, decreasing or constant returns to scale and variable returns to scale. The example outlined above for unit F used constant returns to scale which has a line connecting the leftmost frontier unit to the origin (i.e. the x value of 0). This alternative insists that an optimal production size be determined so that only a minimal input is used to produce the desired output. However, it is often difficult to determine or control the optimal combination of input factors. To remove this uncertain aspect, either constant or decreasing returns to scale or variable returns to scale are used as their efficiency measurements are independent of their production sizes. The efficiency measure based on constant or decreasing returns to scale is determined using the isoquant through points O,A and B (as the line segment intersecting the x-axis must be non-positive). For unit F this value is $\frac{X_{F1}}{X_{BF}}$. The efficiency measure based on variable returns to scale is determined using the isoquant through points A, B and C (as the line segment intersecting the x-axis is unrestricted). For unit F this value is $\frac{X_{F2}}{X_{BF}}$.

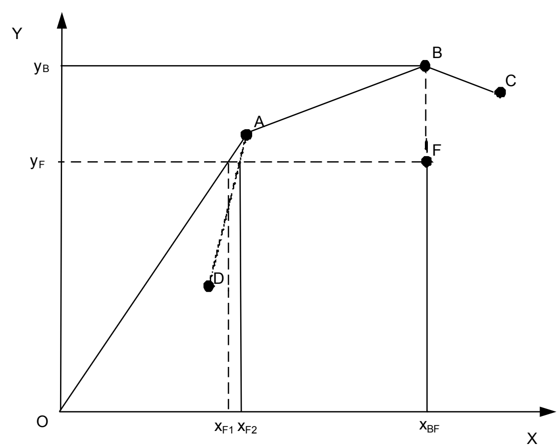

Figure 1: Graphical DEA Representation

There are multiple benefits of using DEA. Most importantly, it relies solely on the empirical data provided and does not depend on a priori assumptions and inferences on the underlying technology or distribution of errors. Being non-parametric, DEA does not involve the assumption that all data can be fitted under a single optimised regression equation which best represents the averages of parameters [Harris II et al., 2000]. Consequently, the results that it produces are more robust than parametric techniques like DFA which are based upon that assumption. It also can incorporate multiple inputs and outputs in the form of weighted virtual values and allows sources of inefficiency to be analyzed and quantified for every evaluated unit. Additionally, DEA does not require a specific production function to be developed (as in Deterministic Frontier Analysis) as all efficiencies on the isoquant have a value of 1. Lastly, the results of DEA focus on relative, not absolute efficiency and identify a realizable benchmark for all studied units. Inefficiencies in sub-optimal centers can be identified in the form of slack in input, output, or both [Lynch et al., 1994]. This slack represents either a surplus or shortage in production, and can be used to identify which parameters contribute the most to the DMU's inefficiency. This offers practical results for policy makers and managers towards which areas to improve in order to become more efficient.

Specifically in regards to healthcare analysis, DEA has become a mainstay since the mid-1980's for measuring productive performance and efficiency, primarily due to its ability to handle multiple inputs and outputs [Hollingsworth, 2003]. According to Hollingsworth (2003), of the 188 studies published on healthcare efficiency between 1988 and 2002, approximately two-thirds utilised DEA, with an additional 20% opting for a two-stage study, identifying efficiency using DEA followed by a form of regression analysis.

The trend towards a more complex DEA analysis is increasing, with only 50% of studies performing a single stage DEA analysis from 1997-2002, down from 60% from 1988-1997. Hollingsworth also suggests the use of secondary regression analysis to further enhance and reinforce findings. However, as a result of the relatively small sample size (n=3) used in this study, only a single stage DEA analysis will be performed.

If the units measured using DEA are representative of a sector (i.e., dialysis treatment centers) the primary disadvantage to using DEA is that it cannot detect an inefficiency of the sector as a whole or define an external standard. In essence, it can only measure how well units perform relative to each other (i.e., that F is less efficient than B in Figure 1) but it cannot determine how efficient frontier units actually are (i.e., if B is perfectly efficient). Another disadvantage is that DEA is sensitive to extreme outlier observations, which can affect the results obtained. Units identified as outliers by other statistical tools may perhaps be judged as really high performing units by DEA, creating an artificially raised efficiency and affecting the overall benchmark.

In general, results from DEA analysis yield important information regarding the optimal operating capabilities of each unit. In addition, by standardizing and weighting inputs, it effectively compares differing entities because it evaluates them according to a given scale. While DEA has proven to be quite a reliable method to determine unit efficiency (having been used to measure efficiency in over 3,200 papers in over 42 countries between 1978-2003 [Emrouznejad et al., in press]) a new method that looks to be very promising in terms of efficiency calculation has recently been developed. This new method is referred to as Free Disposal Hull.

3.4 Free Disposal Hull

Free Disposal Hull (FDH), is a variation of the non-parametric Data Envelopment Analysis technique which assumes free disposability of the production set [Deprins et al., 1984]. While a DEA model is always convex, FDH relaxes this constraint, requiring only that the efficient frontier values are greater than or equal to the actual input and output values. As a result, the FDH model is not restricted solely to a convex model, and provides a more accurate and realistic representation of the efficiency frontier. The FDH model also represents efficiency measures related to an observed production unit, while DEA creates a hypothetical efficient frontier based on weighted inputs and outputs [Hardle et al., 2005].

FDH represents the production feasibility as the union of free disposal hulls of all efficient units. The model uses dominance to select the most efficient units, where a unit is considered dominant if there are no other units that can transform the same amount of input into a greater amount of output. The outcome is that the model produces a production frontier that is realistically attainable, and removes the assumption of convexity. The assumption of convexity also implies that all inputs, outputs, and production activities of a given unit are completely divisible. Thus the non-convex free-disposal hull may better represent situations where large, indivisible capital assets are involved. The non-convexity assumption may prove useful in practical applications, as an actual production unit is more convincing than a hypothetically weighted one.

Drawbacks lie primarily in the accuracy of the results from FDH, which is dependent on the amount of data and inputs and outputs involved. Increasing sample size and insufficient inputs and outputs increases the chance that Decision Making Units are dominated in certain or all categories, resulting in inaccurate results. In addition, because convexity is no longer assumed, the FDH model may not have a feasible solution under a linear program. Due to the low number of Decision Making Units used in this study, and given the tendency of FDH to skew results where sample size is insufficient, FDH will not be used in this analysis.

3.5 Stochastic Frontier Analysis

As mentioned previously, stochastic models differ from deterministic ones in that they may attribute inefficiency to a random element, which is representative of the effects of various external factors. In terms of measuring hospital efficiency, the most commonly used stochastic model is Stochastic Frontier Analysis (SFA), the stochastic version of Deterministic Frontier

Analysis [Hollingsworth, 2003]. Developed in 1977, SFA decomposes error into two distinct parts: cost inefficiency, and a random statistical noise, assumed to be normally distributed [Rosko et al., 2007]. SFA is largely cost oriented, using price data as inputs. Coelli et al. (2005) suggest that while both SFA and DEA each have their own advantages and disadvantages, neither has emerged as the preferred method for efficiency analysis. Rather, each technique should be evaluated on a case-by-case basis. Since the data provided by the Dialysis Registry covered a wide range of inputs, the DEA method is preferred over the SFA method, which requires input to be solely valid price data.

4 Methodology

4.1 Data Collection

Anonymous data was obtained from the 2008 Toronto Dialysis Registry in accordance with Canadian ethical and confidentiality procedures. This data represents the most current and complete available information on dialysis units' workload, organization and processes. As the Toronto Dialysis Registry has the most reliable output data of all Canadian dialysis centers, the study only involved centers in Toronto. Out of scope were dialysis training and home program dialysis units. Due to confidentiality concerns, the identity of each of the three participating centers were not included in this report.

4.2 Production Characteristics

With regards to the data concerning the inputs, preliminary inputs include:

1. Total operating cost
2. Number of nephrologists
3. Number of nurses
4. Number of dialysis technicians
5. Number of pharmacists
6. Number of dieticians
7. Facility size (number of in-center dialysis stations)
8. Number of home hemodialysis patients

9. Number of satellite dialysis patients

10. Number of patients occupying a dialysis station in a day.

The only output to be measured is:

1. Number of dialyzed patients.

4.3 Model

As describe earlier, the efficiency (E), as a ratio of an output (X) and an input (Y), can most simply be expressed as:

$$E = \frac{X}{Y}$$

As DEA incorporates multiple criteria decision based analysis, the multiple outputs and inputs analysed from the system must be transformed into a singular output (x) and input (y) value by weighting their sums respectively. For a given DMU n, a singular output X_n can be represented as:

$$X_n = \sum_i \alpha_i x_{in}$$

and a singular input y_n can be represented as:

$$Y_n = \sum_j \beta_j y_{jn}$$

where α_i is the weight associated with each output i and β_j is the weight associated with each input j. The efficiency equation stated above may now be re-written as:

$$E_n = \frac{x_n}{y_n}$$

Therefore, in order to maximize the efficiency of a given DMU, the input and output weights must be determined such that they maximize the overall efficiency of the system.

The model is constructed such that it maximizes the efficiency of a particular DMU m:

Maximize:

$$E_m = \frac{\sum_i \alpha_i x_{im}}{\sum_j \beta_j y_{jm}}$$

Subject to:

$$\frac{\sum_i \alpha_i x_{in}}{\sum_j \beta_j y_{jn}} \leq 1 \quad \forall\, n \tag{1}$$

$$\alpha_i \beta_j \geq 0 \forall\, i, j$$

To simplify the fractional constraint, it is manipulated in the following manner:

$$\frac{\sum_i \alpha_i x_{im}}{\sum_i \beta_j y_{jm}} \leq 1 \forall\, n \Leftrightarrow \sum_i \alpha_i x_{in} \leq \sum_j \beta_j y_{jn} \,\forall\, n$$

and the following constraint is added to the model:

$$\sum_j \beta_j y_{jm} = 1$$

By doing this, a linear program is created from the original fractional program. The resulting linear program is as follows:

Maximize:

$$E_m = \sum_i \alpha_i x_{im}$$

Subject to:

$$\sum_j \beta_j y_{jm} = 1 \tag{2}$$

$$\sum_i \alpha_i x_{in} \leq \sum_j \beta_j y_{jn} \forall\, n$$

$$\alpha_i \beta_j \geq 0 \,\forall\, i, j$$

To allow the efficiencies to be measured on both a constant and variable returns to scale, a slack variable z is added to the model in the objective function producing the following model:

Maximize:

$$E_m = \sum_i \alpha_i x_{im} + \mathrm{z}$$

subject to:

$$\sum_j \beta_j y_{jm} = 1 \tag{3}$$

$$\sum_i \alpha_i x_{in} \leq \sum_j \beta_j y_{jn} \forall\, n$$

$$\alpha_i \beta_j \geq 0 \,\forall\, i, j$$

4.4 Software

The model was coded using ILog OPL 4.0 software. Microsoft Excel was employed to manipulate the input data. The Solver function was also used to further verify and validate our ILog results.

5 Analysis

To address data variability and the relative sizes of the DMUs, this study analyzed the data both from a constant and a variable returns to scale standpoint.

Table 1 presents the descriptive statistics for the variables used in the efficiency analysis. Many of the variables were consistent with each other, displaying very little variation between centers (i.e., relatively low standard deviations). However, as some variables, such as Number of Home Hemodialysis Patients and Number of Patients Occupying a Dialysis Station in a Day, differed greatly between centers and therefore had comparatively larger standard deviations, both the constant returns to scale (CCR) and the variable returns to scale (VRS) analysis approaches were used and their results compared.

With regards to the uses of each model, as the CCR method measures both technical and scale efficiency, it is used in situations where the DMUs are of relatively equal size. Conversely, the VRS model is used when the DMUs are of notably different sizes, as it measures only technical efficiency. With reference to the model code, the difference between the two models lies in whether the slack variable (i.e., z) is set to zero (as in the CRS model) or is unconstrained (as in the VRS model).

Variables	mean	std. dev.	min	max
Total Operating Cost	24283566	5058281.31	28336065	1861432
Number of nephrologists	7.33333	0.58	8	7
Number of nurses	110.68	45.42	162.96	80.98
Number of dialysis technicians	9.28667	3.82	12.7	5.16
Number of pharmacists	3.1333	0.23	3.4	3
Number of dieticians	3.7	0.61	4.1	3
Number of in-center dialysis stations	47	3.61	51	44
Number of home hemodialysis patients	26.333	33.72	65	3
Number of satellite dialysis patients	14	12.49	24	0
Number of patients occupying a dialysis station in a day	200.67	29.02	229	171
Number of dialyzed patients	415.33	28.57	445	388

Table 1: Descriptive Statistics for Toronto Dialysis Registry

Both models, however, identify which centers perform optimally and sub-optimally by the value of the objective function and the weights of the inputs and outputs. After running each model for each center, and setting the objective function and constraints to optimize the center in question per model run, the optimal efficiency value for that center is determined along with the corresponding input (β_j) and output (α_i) weights. It is important to note that for a center m, the α_i and β_j values constitute the set of the most favourable weights for m, such that the ratio scale is maximized.

A center m is classified as efficient if it has an objective function value (i.e., E_m value) of 1 and there exists at least one optimal input and output weight whose value is greater than 0. If both conditions are not satisfied, the center is classified as inefficient.

The set of all efficient centers are represented in the reference set. For a given run, if a center is efficient, then it constitutes its own reference set. If a center is inefficient, then its reference set is the collection of all centers that are efficient (i.e., that have an efficiency of 1).

Due to the small sample size, the limits of which are further explained in the Discussion section of this report, regression analysis was not performed on the data. Typically, however, either Linear, Censored or Ordinary Least Squares Regression techniques would be employed to identify the significance and correlation of each variable to the overall efficiency. In doing so, the efficiency scores computed from DEA would be used as the dependent variable, while a selection of inputs and outputs would be chosen as the independent variables. The regression model would identify whether a variable is positively or negatively correlated with efficiency, as well as whether the variable was significantly associated to the efficiency calculation. The regression technique selected would depend on, amongst other factors, the nature of the data (i.e., the variable range distribution), the number of dependent variables to be modelled and the complexity of the data. Further information on these commonly used techniques is provided in Sections 5.1 - 5.3.

5.1 Linear Regression Model

The linear regression model is an analysis technique involving one or more independent variables and a dependent variable that changes in response to changes to the independent variables. The variables are modelled using a least squares method, and is represented as a straight line that provides the best fit for the data. While it is a simple model, it is restricted by the number of dependent variables that can be modelled.

5.2 Censored Regression Model

The Censored Regression Model (CRS) is most effectively used in the case where a censored variable is being modelled. Censored variables, typically seen in population statistics, have a large number of observations at a minimum or a maximum, in some cases both. As a result, regression models assuming ordinary sample sizes will show a bias when analyzing censored variables. However, CRS compensates for the censored data by utilizing maximum likelihood estimates to provide a more accurate representation of data behaviour within the regression coefficients [Joreskog, 2002].

5.3 Ordinary Least Squares

Ordinary Least Squares (OLS) is a regression technique used to determine the model that most accurately represents a given set of data. A regression analysis is performed to determine the model with the least amount of squared residuals [Cobb et al., 1928]. The regression analysis is polynomial, but a linear analysis is typically used due to complexity.

Corrected Ordinary Least Squares (COLS) is a variation of the OLS model used in efficiency analysis. The COLS follows the same process as OLS, however the OLS model is then shifted in parallel to pass through the observation with the best value, forming the efficient frontier. The distance between an observation and the efficient frontier is considered the inefficiency of the observation. Due to the parallel shift from the OLS model, the COLS efficiency frontier may produce bias if the original OLS model did not accurately fit the residuals [Aigner et al., 1968].

6 Results

The efficiency scores obtained for all 3 centers within the study using both CRS and VRS are described in Tables 3 and 4, respectively. Each center was determined to be 100% efficient as each had an efficiency score (i.e., an E_m value) of 1. Consequently, each center was the only centre within its reference set. There was little difference between the CRS and VRS models, indicating that the centers were of comparable size and that size was not a large factor in determining the efficiency scores.

Center	A	B	C
Input			
Total Operating Cost ($/year)	18614632	25900000	28336065
Number of nephrologists	7	8	7
Number of nurses	80.98	88.1	135.69
Number of dialysis technicians	5.16	12.7	10
Number of pharmacists	3	3.4	6.51
Number of dieticians	4.1	3	7.3
Number of in-center dialysis stations	46	51	44
Number of home dialysis patients	3	65	11
Number of satellite dialysis patients	18	0	24
Number of patients occupying a dialysis station in a day	171	229	202
Output			
Number of dialyzed patients	445	413	388

Table 2: Input and Output Data

As DEA is used as a benchmarking approach, DMUs are ranked according to efficiency. Table 4 displays the distribution of units within the various

Center	A	B	C
E_m	1	1	1
Reference Set	A	B	C
α_1	0.00224719	0.0024213	0.00257732
β_1	0.00000295	0.00000123	0.00000122
β_2	0.19699894	0.20684021	0.20687518
β_3	0	0	0
β_4	0	0	0
β_5	0	0	0
β_6	0	0	0
β_7	0	0	0
β_8	0	0	0
β_9	0	0.02096999	0.02097757
β_10	0	0	0
z	-	-	-

Table 3: CRS Efficiency Scores and Weights

ranked categories. As each center had an efficiency of 1, no comparison may be made between centers as each is ranked as having 100% efficiency.

Center	A	B	C
E_m	1	1	1
Reference Set	A	B	C
α_1	0.00224719	0.0024213	0.00257732
β_1	0.00000295	0	0
β_2	0.19699894	0.00083364	0.00086861
β_3	0	0.00768805	0.00753678
β_4	0	0	0
β_5	0	0.00040767	0.00038727
β_6	0	0.00059225	0.0005787
β_7	0	0.00391253	0.00416081
β_8	0	0	0
β_9	0	0.00459265	0.00460022
$\beta_1 0$	0	0.01153051	0.01210831
z	0	0.00005001	0.00005211

Table 4: VRS Efficiency Scores and Weights

	CRS Model	VRS Model
Dialysis Units (n)	3	3
Mean	1	1
Standard Deviation	0	0
Median	1	1
Minimum	1	1
Maximum	1	1
Ranking		
100%	3	3
<99%	-	-

Table 5: Ranked Efficiencies

7 Discussion

While it could simply be concluded that all 3 dialysis centers were efficient, studies have shown that having a small sample size in DEA runs the risk of having a larger percentage of DMU's in the dominant set (i.e. the set of DMU's that have an E_m of 1). Alirezaee et al. (1988) suggest that a smaller sample size coupled with a larger percentage of DMU's in the dominant set creates a bias in efficiency evaluation. The degree of bias is based on the degree of the two factors previously mentioned. For example, in the study of banks, a sample size of 20, with approximately 55% of the sample in the dominant set was shown to have a biased average efficiency score that was roughly 60% higher than its true average efficiency score. Since the sample size for this paper was only 3 and 100% of DMU's were in the dominant set, it is safe to assume the efficiencies calculated are much higher than they truly are. Alirezaee also suggests that this bias is further increased with larger numbers of inputs and outputs. Therefore, in order to mitigate the bias in efficiency evaluation, the DEA study on dialysis centers must be performed on a much larger sample size.

To demonstrate the value of DEA on a greater sample size, mock centers D,E,F,G,H and I were added with input variables identical to those of center A and the outputs as described in Table 6.

Table 7 displays the results obtained from running the CRS model with the

Unit	Output
A	445
B	413
C	388
D	540
E	440
F	403
G	350
H	397
I	260

Table 6: Ranked Efficiencies

mock data. As described, increasing the sample size decreased the number of variables that attained a 100% efficiency rating. It also provides greater granularity into the relative centre ratings and therefore provides greater use as a benchmarking tool.

	CRS Model
Dialysis Units (n)	9
Mean	0.80555
Standard Deviation	0.17752
Median	0.81481
Minimum	0.48148
Maximum	1
Ranking	
100%	3
90-99%	0
80-89%	2
70-79%	2
60-69%	1
50-59%	0
<50%	1

Table 7: Ranked Efficiencies

8 Conclusion

With further data, Data Envelopment Analysis should prove useful in providing an empirical, non-assumptive benchmarking analysis of Toronto dialysis centers. Until such data is obtained, however, suboptimal centers cannot be identified and no definite recommendations may be made to improve the efficiencies of these centres.

References

DJ Aigner and S. Chu, "On Estimating the Industry Production Function", American Economic Review 58, 826-839 (1968).

M Alirezaee and M Howland, "Sampling Size and Efficiency Bias in Data Envelopment Analysis", Journal of Applied Mathematics Decision Sciences 2(1), 51-64 (1998).

SV Birman and PE Pirondi and EY Rodin, "Application of DEA to medical clinics", Mathematical and Computer Modelling 37, 923-936 (2003).

WF Bowlin, "Measuring Performance: An introduction to Data Envelopment Analysis (DEA)", Journal of Cost Analysis 3-27, (1998).

Canadian Institute for Health Information, "National Health Expenditure Trends", (1975-2005).

Charnes and WW Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units", European Journal of Operations Research, 429-444 (1978).

CW Cobb and PH Douglas, "A Theory of Production", American Economic Review 18, 139-165 (1928).

T. Coelli and D. Rao and C. O'Donnell and G. Battese, "An Introduction to Efficiency and Productivity Analysis (2nd ed)",New York: Springer, (2005).

D. Deprins and L. Simar and H. Tulkens, "Measuring labor inefficiency in post offices", in: M. M. P. Pestieau and H. Tulkens (Eds.), "The Performance of Public Enterprises: Concepts and Measurements", North-Holland, Amsterdam,243267, (1984).

A. Emrouznejad and B. Parker and G. Tavares, "A bibliography of Data Envelopment Analysis (1978-2001)", Sic-Econ Plan Sci (in press).

MJ Farrell, "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society 120, 253-290 (1957).

Karen Gerard and Paul Roderick, "Comparison of Apparent Efficiency of Haemodialysis Satellite units in England and Wales using Data Envelopment Analysis", International Journal of Technology Assessment in Healthcare 19, 3 (2003).

Dimitris I Giokas, "Greek Hospitals: How Well their Resources are used", Omega 29, 73-83 (2001).

Government of Canada, "Diabetes: An Investment for the Future Health of Canadians", Association presentation to the House of Commons Standing Committe on Health, (October 2003).

Wolfgang Hardle and Seok-Oh Jeong, "Nonparametric Productivity Analysis", March 2005.

J. Harris II and H. Ozgen and Y. Ozcan, "Do mergers enhance the performance of hospital efficiency?", Journal of the Operational Research Society 51, 801-811 (2000).

Bruce Hollingsworth, "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care", Health Care Management Science 6:44, 203-218, (2003).

KG Joreskog, "Censored Variables and Censored Regression", Econometric Theory, (2002).

Maria Katharaki, "Approaching the management of hospital units with an operation research technique: The case of 32 Greek obstetric and gynaecology public units", Health Policy 85, 19-31 (2005).

Nick Kontodimopoulos and Dimitris Niakas, "Efficiency Measurement of Hemodialysis units in Greece with Data Envelopment Analysis", Health Policy 71, 195-204 (2005).

Peter Kooreman,"Nursing Home Care in The Netherlands: a nonparametric efficiency analysis", Journal of Health Economics 13, 301-316 (1994).

Timo Kuosmanen, "Data Envelopment Analysis as Least-Squares Regression", MTT Dicussion Paper 3/2006.

J. Lynch and Y. Ozcan, "Hospital Closure: An Efficiency Analysis", Hospital and Health Services Administration (Chicago) 39-2, 205 (1994).

JA Nyman and DL Bricker, "Profit Incentives and Technical Efficiency in the Production of Nursing Home Care", The Review of Economics and Statistics 56, 586-594 (1989).

Provincial and Territorial Ministers of Health, "Understanding Canada's Health Care Costs, Final Report" (2000).

M. Rosko and R. Mutter, "Stochastic Frontier Analysis of Hospital Inefficiency", Medical Care Research and Review 20-10, (2007).

Paul Smith, "Model Misspecification in Data Envelopment Analysis", Annals of Operations Research 73, 233-252, (1997).

Appendix A: Accountability

Report Sections:

Abstract - Sara Dolcetti and Stephen Chu

1.0 Introduction - Sara Dolcetti

2.0 Dialysis Treatment in Canada - Sara Dolcetti

3.0 Efficiency Measurement - Sara Dolcetti

3.1 Ratio Analysis - Stephen Chu

3.2 Deterministic Frontier Analysis - Stephen Chu

3.3 DEA - Sara Dolcetti

3.4 Free Disposal Hull - Stephen Chu

3.5 Stochastic Frontier Analysis - Stephen Chu

4.0 Methodology 4.1 Data Collection - Sara Dolcetti and Stephen Chu

4.2 Production Characteristics - Sara Dolcetti and Stephen Chu

4.3 Model - Sara dolcetti

4.4 Software - Sara Dolcetti and Stephen Chu

5.0 Analysis - Sara Dolcetti

6.0 Results - Sara Dolcetti

7.0 Discussion - Sara Dolcetti and Stephen Chu

8.0 Conclusion - Sara Dolcetti

Bibliography - Stephen Chu

Citations - Stephen Chu

Model and Software:

Formatting and Latex Implementation - Stephen Chu

Model Formulation and Coding in Excel and Ilog - Sara Dolcetti

Excel Data Manipulation, Descriptive Statistics - Stephen Chu