Measuring Routine Nursing Service Efficiency: A Comparison of Cost per Patient Day and Data Envelopment Analysis Models

Thomas R. Nunamaker

This article provides an illustrative application of Data Envelopment Analysis (DEA) methodology to the measurement of routine nursing service efficiency at a group of Wisconsin hospitals. The DEA efficiency ratings and cost savings estimates are then compared to those resulting from application of Medicare's routine cost limitation to the sample data. DEA is also used to determine if any changes in the potential for efficient operations occurred during the 1978-1979 period.

Empirical results were representative of the fundamental differences existing between the DEA and cost per patient day approaches. No evidence was found to support the notion that the overall potential for efficient delivery of routine services by the sample institutions was greater in one year than another.

INTRODUCTION

An ongoing problem in virtually all nonprofit organizations concerns the development of useful efficiency measures. In general, the problems involve the absence of an acceptable summary statistic, i.e., difficulties with measuring and combining multiple outputs and relating them to the

---

This research was supported by a grant from the W.K. Kellogg Foundation through The Center for Advanced Study in Health Care Fiscal Management, Organization, and Control at the University of Wisconsin-Madison. Thomas R. Nunamaker is Assistant Professor of Accounting and Business Law, Washington State University, Pullman WA 99164. Address communications and requests for reprints to him at that address.

The article is followed by an invited commentary by Arie Y. Lewin, Professor of Business Administration at Fuqua School of Business at Duke University. For communications and/or requests for reprints of his commentary, contact Dr. Lewin at Fuqua School of Business, Duke University, Durham NC 27706.

resources consumed in producing those outputs. One aspect of the problem is that most nonprofit organizations do not operate in competitive markets, implying net income (or profit) does not provide a useful indicator of productive efficiency [1, 2].

In profit-making firms, the profit figure does represent one measure of efficiency and effectiveness, even in the presence of multiple objectives [3]. According to the economic theory of the firm, in perfectly competitive markets, prices charged by different firms for a given common output are the same. Further, since prices reflect aggregate market supply and demand conditions, such prices cannot be influenced by any single firm and are representative of the opportunity cost of productive inputs as well as the consumption choices of individual consumers. Hence, in perfectly competitive markets, supply and demand conditions will force firms to charge a unique optimum price for similar outputs, so that price differences generated by inefficiency will fail to persist. Given this scenario, profits do provide a useful indicator of output/input efficiency. Observe that relaxing the perfectly competitive markets condition does not substantially alter this conclusion. The existence of some market pressures will maintain profits as an efficiency standard for profit-oriented entities.

Such is not the case in not-for-profit oganizations which are typically insulated from most competitive market forces.[1] As Lewin and Morey [4] state:

> While it is true that private firms pursue multiple goals and that goals often are not sufficiently well defined, the market system does provide economic (as distinct from social) indicators of performance through such measures as profits, rates of return on investments, market share, and so forth. The public sector lacks both an analog for profit seeking behavior and an adequate feedback system for learning about the quality of decisions. As a result the problem of evaluating performance of public sector organizations and the development of insights to guide performance improvement has been much more difficult. (p.1)

However, the Data Envelopment Analysis (DEA) methodology pioneered by Charnes, Cooper and Rhodes [5] would seem to provide a potentially valuable model for judging the relative efficiency of nonprofit organizations performing like tasks. DEA utilizes a fractional linear programming technique to overcome traditional performance measurement problems in nonprofits by (1) being capable of handling non-commensurate multiple inputs and outputs, (2) not being dependent on any a priori arbitrary weighting scheme, and (3) providing a single summary measure of the relative efficiency of a nonprofit firm compared with all other such firms producing similar outputs. Note DEA does not

address the problem of individual output measurement (e.g., patient days versus improvement in patient's health). Instead, given a particular set of multiple input and output measurements, DEA seeks to combine them in a nonarbitrary, nonsubjective fashion via the criteria of Pareto Efficiency.

The purpose of this paper is to apply the DEA methodology to an oft-alleged inefficient sector of the economy: nonprofit hospitals. However, due to the heterogeneous and ill-defined nature of total hospital output (see [6]), the present study will focus only on inpatient routine nursing services provided by a sample of Wisconsin hospitals. This limited scope approach should permit investigation of the usefulness of DEA in a hospital setting while reducing the study to manageable proportions.

We might observe that determination of the relative efficiency of hospitals in providing routine patient care is in itself a significant problem. The importance of this issue is reflected in Section 223 of Public Law 92–603, which empowers the Secretary of HEW (now HHS) to limit Medicare reimbursement to the "necessary costs of *efficiently* delivering covered services" (emphasis added) [7]. HEW's initial attempt to limit Medicare payments to hospitals came in 1974 with publication of regulations establishing reimbursement limits for inpatient routine service costs. The specific limits established were based on the Section 223 laws:

> In the determination of allowability of provider costs, the costs determined to be in excess of those necessary in the *efficient* delivery of needed health services are excluded. ([8] Section 2500, Rev 129, emphasis added).

Presumably, HEW's reimbursement limits reflect the cost an efficient hospital incurs in providing a day of routine service to Medicare patients. Hospitals with a routine cost per patient day exceeding the limit are assumed inefficient while providers falling below the limit are deemed efficient. The original limits effective for cost reporting periods beginning after June 1974, were set in the following manner:

> The basic data for setting prospective limits for hospitals were derived from the average per diem costs for general inpatient routine services for each participating short-term hospital. . . Each hospital was placed in one of the 70 classification groups based on location and bed size (see § 2510.1) and the short-term hospital data were arrayed for each class. The 90th percentile and the median for each class were determined. Ten percent of the median for each class were added to the 90th percentile amount for that class, thereby establishing the general inpatient routine cost ceiling for each class. ([8] Section 2520.2, Rev. 159).

Although the limit determination procedure has been revised several times over the past few years,[2] its arbitrary nature persists, since for

example, little empirical evidence is provided to support setting limits at the 90th percentile. Furthermore, since a single cost per diem statistic serves as the Medicare efficiency measure, it appears no weights have been applied to the different types of routine patient days (e.g., aged, maternity, etc.) to reflect differences in output "value" and mix. The present study will attempt to overcome these difficulties via the DEA model, by constructing a piece-wise linear efficient production function from the subset of sample hospitals, and comparing the remaining hospitals to this efficient surface. Additionally, the weights accorded each output will be objectively determined from the data in accordance with Pareto Efficiency conditions.

Comparisons of efficiency ratings as well as the total cost savings possible under each model (DEA vs. cost/patient day) should provide some insights into the adequacy of Medicare's efficiency measurement and limit-setting procedures. Since DEA represents a decision model using accounting information inputs, another important question concerns the model's sensitivity analysis on DEA data inputs for selected sample institutions.

The paper is organized as follows. An overview of the nonlinear DEA model along with its linear programming equivalents is presented in the next section. The operational efficiency evaluation model applicable to hospital routine nursing services is developed in the third section, data analysis and discussion of results is provided in the fourth, and some brief concluding remarks are presented in the fifth, and final, section.

OVERVIEW OF THE DEA MODEL

DEA is a technique for evaluating the relative efficiency of "decision-making units" (DMUs) producing similar outputs. It represents an extremal predictor in that the most efficient DMUs within a group serve to define a piece-wise linear production function. The remaining DMUs are then evaluated relative to this efficient surface. Hence, the efficiency criteria employed is maximization of relative efficiency rather than engineering efficiency. Due to the absence of competitive output markets for nonprofits and difficulties specifying technical capacities through engineering studies, the reference set used in the proposed evaluation scheme must necessarily depend upon the most efficient DMUs observed.[3]

Conceptually, the DEA model represents a fractional programming approach to efficiency evaluation for DMUs having multiple inputs and outputs. The problem is formulated so as to select weights (u_r and v_i) which maximize the ratio of a linear combination of outputs (y_{rj}) to a

linear combination of inputs (x_{ij}). The scalar efficiency measure (h_o) for DMU_{j_o} (where o denotes the DMU being evaluated) is constrained in that the weights selected must be feasible and cannot result in an efficiency ratio greater than that observed for the best DMU in the group. Mathematically, the model appears as follows:

$$\text{Maximize} \quad h_o = \frac{\sum_{r=1}^{s} u_r y_{rj_o}}{\sum_{i=1}^{m} v_i x_{ij_o}} \tag{1}$$

$$\text{Subject to:} \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \,,\; j = 1,2,\ldots,n \text{ DMUs}$$

$$u_r > 0 \,,\; r = 1,2,\ldots,s \text{ outputs}$$

$$v_i > 0 \,,\; i = 1,2,\ldots,m \text{ inputs}$$

Note the output/input ratio for DMU_{j_o} contained in the functional is also included as one of the $j = 1,2,\ldots,n$ constraints, thus assuring the best efficiency ratio is ≤ 1.0. Essentially, the model requires comparing all DMUs to find the most efficient subset of DMUs to form the reference group against which DMU_{j_o} is evaluated.

The notion of efficiency employed in this approach is termed Pareto Efficiency (or Pareto-Koopmans Efficiency) which in turn is a natural extension of the well-known social choice criterion of Pareto Optimality. Pareto Efficiency may be paraphrased from Bessent and Bessent [11] as follows:

> A DMU is not efficient in producing its output (from given amounts of input) if it can be shown that some other DMU or combination of DMUs can produce more of some output, without producing less of any other output and without utilizing more of any resource. Conversely, a DMU is efficient if this is not possible. (2)

Note the above definition reflects an output orientation. Pareto Efficiency may similarly be characterized with an input orientation:

> A DMU is not efficient in utilizing its inputs (to produce given amounts of output) if it can be shown that some other DMU or

combination of DMUs can produce the same amount of output with less of some resource and no more of any other resource. Conversely, a DMU is efficient if this is not possible. (3)

Evaluation of the nonlinear form in (1) is not a trivial task. However, Charnes, Cooper and Rhodes (hereafter referred to as CCR) [5] provide a practical solution to this difficulty by providing transformations which reduce (1) to the following ordinary linear programming problem:

$$\text{Minimize} \quad h_o - \epsilon \sum_{i=1}^{m} s^{-\prime}_{ij_o} - \epsilon \sum_{r=1}^{s} s^{+\prime}_{rj_o} \tag{4.1}$$

$$\text{Subject to: } - \sum_{j=1}^{n} x_{ij} \lambda'_j + x_{ij_o} h_o - s^{-\prime}_{ij_o} = 0, \qquad i = 1, 2, \ldots, m \text{ inputs}$$

$$\sum_{j=1}^{n} y_{rj} \lambda'_j - s^{+\prime}_{ij_o} = y_{rj_o} \quad , \qquad r = 1, 2, \ldots, s \text{ outputs}$$

$$\lambda'_j \geq 0 \quad , \qquad j = 1, 2, \ldots n \text{ DMUs}$$

$$s^{-\prime}_{ij_o} \text{ (input slack)} \geq 0 \quad , \qquad i = 1, 2, \ldots, m$$

$$s^{+\prime}_{rj_o} \text{ (output slack)} \geq 0 \ , \qquad r = 1, 2, \ldots, s$$

h_o unrestricted in sign

Alternatively, by considering the inefficiency form of (1), i.e., inputs divided by outputs, we can obtain the solution to (1) and hence (4.1) by solving the following problem:

$$\text{Maximize} \quad z_o + \epsilon \sum_{r=1}^{s} s^{+}_{rj_o} + \epsilon \sum_{i=1}^{m} s^{-}_{ij_o} \tag{4.2}$$

$$\text{Subject to: } - \sum_{j=1}^{n} y_{rj} \lambda_j + y_{rj_o} z_o + s_{rj_o} = 0, \qquad r = 1, 2, \ldots, s \text{ outputs}$$

$$\sum_{j=1}^{n} x_{ij} \lambda_j + s^{-}_{ij_o} = x_{ij_o} \quad , \qquad i = 1, 2, \ldots, m \text{ inputs}$$

$$\lambda_j \geq 0 \quad , \qquad j = 1, 2, \ldots n \text{ DMUs}$$

$s^{+}_{rj_o}$ (output slack) ≥ 0 , $r = 1, 2, \ldots, s$

$s^{-}_{ij_o}$ (input slack) ≥ 0 , $i = 1, 2, \ldots, m$

z_o unrestricted in sign

where ϵ in (4.1) and (4.2) is an infinetesimal value (e.g., 10^{-6}) which guarantees the strict positivity of u_r and v_i in (1) without disturbing the optimal solution to the problem (see [5, 12]). Note that $h_o^* = 1/z_o^*$ where * indicates an optimal solution value. Hence, z_o^* can viewed as an "inefficiency" score (inputs divided by outputs) analogous to the cost/patient day measure.

The h_o^* score having $1 \geq h_o^* > 0$, where $h_o^* = 1$ indicates an efficient DMU, can be interpreted as a scalar efficiency measure according to a "proportionate contraction of resources" criteria. That is, $(1 - h_o^*)$ measures the reduction in each input required for the DMU being evaluated to produce its given outputs as efficiently as the subset of DMUs to which it is being compared. Note that if inputs are measured in cost terms, it becomes possible to calculate the total cost savings assuming inefficient hospitals (as determined via (4.1) or (4.2)) were to become efficient.[4]

Finally, we should make explicit two assumptions of the model, which may already be apparent. First, the DEA approach assumes constant returns to scale for each DMU evaluated. Second, it is assumed the constructed efficient production frontier is piece-wise linear and continuous. Importantly, this second assumption implies all points along the efficient surface are practically attainable production possibilities. Together, these two assumptions permit h_o^* to be interpreted as a scalar efficiency measure.

This concludes our brief discussion of the DEA model.[5] In the next section, we will adapt the model for efficiency evaluation of hospital routine nursing services.

Routine Nursing Service Efficiency Measurement

SAMPLE SELECTION

Since DEA measures the relative efficiency of DMUs performing similar tasks, the sample hospitals selected should be fairly homogeneous with respect to product output. Within the DEA framework, it makes little sense to compare a large teaching institution to a small community

hospital; the goals and hence the healthcare outputs of these providers are radically different. These output differences are then reflected in significantly different organizational structures and cost functions.

Numerous models are available for grouping hospitals in order to achieve intragroup homogeneity, and substantial debate surrounds the desirability of one model versus another.[6] The present study will not address the hospital classification issue. Instead, sample hospitals were selected from a single cluster defined by the Wisconsin Hospital Clustering Model, and given this model, the sample hospitals are assumed fairly homogeneous along output dimensions.[7] Reliance on the Wisconsin classification model also seems reasonable from a practical viewpoint, since the resultant clusters are used by the Wisconsin Rate Reveiw Program to develop peer group standards against which the performance of hospitals in each cluster are evaluated.

The largest cluster available ($n = 17$) was chosen from the 1978 Wisconsin Hospital Groupings. In general, the sample institutions were small in size (average number of beds = 118) with little or no educational activities indicated, thus reducing measurement problems associated with teaching outputs.

VARIABLE SELECTION

As noted previously, a primary measure of hospital efficiency used by the Federal Government for health insurance reimbursement purposes is the cost/patient day of general inpatient routine nursing care. General routine nursing care consists of those services included in the daily service charge. Included in general routine services are the regular room (excluding special care units), dietary and nursing services, minor medical and surgical supplies, medical social services, psychiatric social services, and the use of certain equipment and facilities for which a separate charge is not customarily made.

In essence, Medicare's efficiency measure consists of multiple inputs (e.g., direct nursing hours, supplies, allocated overhead, etc.) aggregated by price (cost) weights, and the total divided by an equally weighted measure of outputs (patient days). Certainly, a strong argument can be made that the input weights (historical costs) are not relevant or proper for reimbursement and efficiency measurement purposes. However, this difficulty could be resolved by applying general and/or specific price level adjustments. At least conceptually, then, aggregation of hospital inputs via price weights would appear acceptable.

Routine outputs present a much more difficult problem since no objective market price weights exist which can serve to aggregate dissimilar outputs into a single summary statistic. Hence, a common

solution to the problem (that taken by the Medicare program) is to weight each routine output equally, and simply add together all the different types of routine patient days. Such an approach ignores any resource consumption differences existing among different classes of routine patients. Some acknowledgment of these differences by the Medicare program is evidenced by the provision of an 8½% nursing salary differential applicable to all aged, pediatric, and maternity patients.[8] Even in this provision, however, a degree of arbitrariness exists since aged, pediatric, and maternity days are weighted equally.

The present study focuses on objective determination of weights for given routine outputs, while accepting total inpatient costs as a satisfactory aggregation of multiple routine inputs. The results, though, should be viewed in light of previous caveats concerning the usefulness of historical cost weights. At the same time, however, consideration of a single aggregated input variable permits hospitals the widest latitude for taking action to improve their efficiency rating. That is, the "do it any way you can" approach (apparently) adopted by the Medicare program to control routine service costs, is preserved while explicitly taking into account the varying mix of routine outputs which exists across hospitals.

Two years of data (1978, 1979) were obtained for the 17 sample hospitals. One may be tempted to combine observations from both years, resulting in 34 DMUs. Applying DEA in this way, however, requires assuming that the production possibilities present in 1978 were also present in 1979, and that no significant overall constraints to efficient operations distinguished one year from the other.

We can avoid this assumption by applying the techniques described in CCR [10] and treat each year as a discrete production period or "program." In Phase One of the analysis, DEA is used to estimate separate production frontiers for the 1978 versus 1979 observations. Employing the efficiency adjustments detailed in [5], the inputs and outputs of each DMU are simultaneously adjusted so that all DMUs are positioned on their respective production surface (1978 vs. 1979).

In Phase Two, the adjusted data for the two years is combined into a single sample, and the model rerun, to ascertain any changes in efficient production possibilities that may have occurred over the 1978–1979 period. These procedures, in effect, permit us to disentangle "managerial" from "program" efficiencies where the programs we refer to are discrete time intervals. Consequently, in evaluating each hospital's efficiency, we can control for constraining events and forces that may have had an overall impact on one year and not the other.

For the 17 sample hospitals, total inpatient routine service costs, total aged, pediatric, and maternity patient days, and total all other patient days were obtained from 1978 and 1979 Medicare cost reports. To remove

Table 1

Input Variable	*Output Variables*
Run A: 16 DMUs	
1. Total inpatient routine costs	1. Total routine aged and pediatric days 2. Total routine maternity days 3. All other routine days
Run B: 11 DMUs	
1. Total inpatient routine costs	1. Total routine aged and pediatric days 2. Total routine maternity days 3. All other routine days
Run C: 11 DMUs	
1. Total inpatient routine costs	1. Total routine aged days 2. Total routine pediatric days 3. Total routine maternity days 4. All other routine days

inflationary effects, the 1979 cost data were deflated (in the aggregate) to 1978 dollars using the Hospital Cost Index. For 16 hospitals (32 DMUs), it was possible to identify separately total maternity patient days via AHA Annual Survey. Total pediatric patient days could be identified separately for 22 of those 32 DMUs. Thus, in Phase One, three separate DEA runs employing different sample sizes and/or output variable specifications were examined for each time period under consideration (see Table 1). Phase Two consists of comparing the efficient frontiers attained in Phase One for 1978 and 1979, after all managerial inefficiencies have been eliminated. Hereafter, each hospital will be referred to as H1 through H16. Additionally, each run in Table 1 will be referred to as *Run* A_t through C_t where $t = 1,2$ corresponds to 1978 and 1979 observations respectively. For instance, *Run* A_2 refers to DEA using *Run A* variables applied to 1979 observations.

RESULTS AND DISCUSSION

PHASE ONE: EVALUATION OF MANAGERIAL EFFICIENCY

A comparison of h_o^* scores from *Run* A_1 (1978) and *Run* A_2 (1979) with the routine cost/patient day calculated for each DMU is provided in Table 2.

Table 2: Comparison of DEA *Run* A_t and *Cost/Patient Day* Efficiency Measures—Phase One
(Efficiency Rankings in Parentheses: *1* = most efficient)

	1978		*1979*	
Hospital	h_o^* *From Run* A_1	*Routine Cost/Patient Day*	h_o^* *From Run* A_2	*Routine Cost/Patient Day*
H1	.99 (5)	$ 77.91 (4)	1.00 (1)	$ 77.55 (5)
H2	.91 (9.5)	81.89 (7)	.85 (11)	86.95 (10)
H3	.88 (12)	90.66 (12)	.86 (9.5)	86.73 (9)
H4	1.00 (1)	72.94 (2)	1.00 (1)	75.74 (4)
H5	.73 (15)	94.33 (13)	.75 (15)	99.13 (14)
H6	.97 (6.5)	74.46 (3)	1.00 (1)	75.31 (3)
H7	1.00 (1)	67.54 (1)	1.00 (1)	72.66 (1)
H8	.89 (11)	99.98 (15)	.83 (12)	106.96 (16)
H9	.84 (13)	86.42 (10)	.88 (8)	85.55 (8)
H10	.72 (16)	100.13 (16)	.71 (16)	101.99 (15)
H11	.97 (6.5)	87.52 (11)	.86 (9.5)	90.53 (11)
H12	.95 (8)	80.62 (6)	.80 (13.5)	91.08 (12)
H13	.91 (9.5)	78.48 (5)	.91 (7)	82.62 (7)
H14	.82 (14)	94.64 (14)	.80 (13.5)	95.62 (13)
H15	1.00 (1)	85.08 (9)	1.00 (1)	81.85 (6)
H16	1.00 (1)	81.95 (8)	1.00 (1)	73.25 (2)
Correlation Between *Cost/Patient Day* and h_o^* Efficiency Rankings		*1978* 73	*1979* .94	

(Subsequent discussion will focus upon Runs A_1 and A_2, although similar comments will apply to Runs B_t and C_t when compared to the cost/patient day results.)

Inspection of Table 2 reveals several interesting facts. First, a strong negative association exists between the h_o^* and cost/patient day measures. This suggests large values of h_o^* (indicating high efficiency) tend to be paired with small values for the cost/patient day measure (also indicating high efficiency) and vice versa (correlation based on 1978 observations = −.76, 1979 observations = −.90).

Second, a high correlation (.86) exists between the 1978 and 1979 h_o^* scores. Hospitals that exhibited high (low) managerial efficiency during 1978 continued to do so in 1979.

Third, note that although the efficiency rankings under the two models are similar, some important differences exist. Consider for instance H15, which under *Run* A_1 has $h_o^* = 1$ yet a cost/patient day slightly higher than the median of $83.52/day. This occurs because H15 Pareto dominates all other DMUs in producing "other routine patient days." Its relatively high cost/patient day is permitted within the Pareto concept since it produces a fundamentally different mix of outputs.

This observation underscores a strength of DEA particularly in relation to hospital efficiency measurement. That is, DMU_{j_o} is compared to those Pareto Efficient DMUs which are, loosely speaking, most similar to DMU_{j_o} in terms of input and output mix (see [5]). In this study, since only one input variable is considered, a particular hospital's h_o^* score is generated by reference to those Pareto Efficient DMUs which are "most similar" to the evaluated hospital such that a constant output mix is maintained. Hence, inefficient DMUs are not being compared to points on the efficient production frontier having a different output mix (implying different resource consumption requirements). This is a desirable feature of the model since output mix seemingly is determined by factors (e.g., population demographics, mix of physician specialties, physician practice characteristics, etc.) largely not under the control of hospital management, at least in the short run. Hence, management of H15 is not penalized under DEA for producing a higher cost output mix which differs from that produced by other institutions in the sample.[9]

The previous discussion has relied on correlation analysis and comparison of rankings in order to uncover differences between the DEA and cost/patient day models. From a regulator's perspective, however, such analysis is probably not very insightful in terms of setting reimbursement limits and controlling costs. Of perhaps greater interest, is the potential cost savings suggested by the DEA model when compared with that resulting from application of Medicare's routine limit criteria to the sample institutions.

The total potential routine cost savings for each DEA Run (A_t–C_t, $t=1,2$) was calculated by multiplying DMU_j's total routine costs ($j=1,2,\ldots,16$ for Run A_t; $j=1,2,\ldots,11$ for Runs B_t and C_t) by its corresponding ($1-h_o^*$) value; and summing across DMUs. In determining the cost savings possible using the Medicare criteria, the particular criteria in effect during 1978 and 1979 were used. The routine cost/patient day for all DMUs in a given year were placed in increasing order and the 80th percentile determined. To this figure, 10% of the median was added to arrive at the routine limit for the DMUs under examination. These results are summarized in Table 3.

The findings in Table 3 are quite striking.[10] Under the DEA model, we find the average cost savings per inefficient DMU is several times larger than that resulting from application of the Medicare criteria to the sample institutions. Also provided in Table 3 is a comparison of the total potential cost savings under each model for 1978 and 1979. This comparison, however, will be biased in favor of the DEA model since all hospitals not on the efficient production frontier are classified as inefficient, while the Medicare model identifies only the extreme outliers as inefficient

Table 3: Potential Cost Savings—Summary of Results (Phase One)

	1978				*1979*			
	Run A_1	*Run B_1*	*Run C_1*	*Medicare Criteria*	*Run A_2*	*Run B_2*	*Run C_2*	*Medicare Criteria*
Number of DMUs	16	11	11	16	16	11	11	16
Average h_o^*:								
All inefficient DMUs	.88	.85	.87	—	.83	.83	.84	—
All DMUs	.91	.91	.92	—	.89	.89	.90	—
Number (%) of DMUs which are inefficient	4 (25%)	4 (36%)	4 (36%)	0 (0.0%)	6 (37.5%)	4 (36%)	4 (36%)	1 (6.3%)
Average potential cost savings per inefficient DMU	$ 336,276	$ 422,258	$ 369,929	$ 0	$ 480,474	$ 519,191	$ 469,097	$ 101,220
Total Potential cost savings	$4,035,309	$2,955,809	$2,589,502	$ 0	$4,804,735	$3,634,342	$3,283,681	$ 101,220

DMUs. As a consequence, none of the 1978 DMUs and only one from 1979 are inefficient according to the Medicare criteria. DEA under *Run* A_t, on the other hand, identifies 4 and 6 inefficient DMUs during 1978 and 1979 respectively. The DEA results from *Run A* imply a total potential cost savings for all 16 DMUs of $4,035,309 in 1978 and $4,804,735 in 1979 compared with zero in 1978 and $101,220 in 1979 per the Medicare criteria. Similar observations apply to cost savings comparisons of DEA Runs B_t and C_t with Medicare criteria when both are applied to the DMUs in these runs.

Finally, comparison of h_o^* scores across Runs A_t–C_t for a given year provides some insight into the importance of sample selection and/or variable specification when using DEA (see Table 4).

Observe that for some DMUs in a given year, h_o^* remains fairly constant across all three runs. For several DMUs, however, a decreased sample size and/or a disaggregation of output variable results in a higher h_o^*. In *no* instance does the reduction in sample size (Runs A_t to B_t) decrease h_o^* for a particular DMU. Intuitively, the impact of a decreased sample size on these findings is not unexpected. Removing DMUs from a sample can do no worse than leave h_o^* for the remaining DMUs unchanged. Furthermore, when an efficient DMU (e.g., H15) is removed, then a previously inefficient DMU (e.g., H8) may now become Pareto Efficient, which in turn may increase (but not decrease) h_o^* for the other DMUs in the run.

Table 4: Comparison of h_o^* from DEA Runs A_t, B_t, C_t (Phase One)

	1978 h_o^* from Runs			*1979 h_o^* from Runs*		
Hospital	A_1	B_1	C_1	A_2	B_2	C_2
H1	.99	—	—	1.00	—	—
H2	.91	.91	.97	.85	.85	.87
H3	.88	.88	.92	.86	.92	.91
H4	1.00	1.00	1.00	1.00	1.00	1.00
H5	.73	.73	.75	.75	.75	.77
H6	.97	.97	.96	1.00	—	—
H7	1.00	1.00	1.00	1.00	1.00	1.00
H8	.89	1.00	1.00	.83	1.00	1.00
H9	.84	—	—	.88	.88	.99
H10	.72	.72	.72	.71	.71	.72
H11	.97	—	—	.86	—	—
H12	.95	.95	.95	.80	.81	.81
H13	.91	—	—	.91	—	—
H14	.82	.82	.84	.80	.80	.81
H15	1.00	—	—	1.00	—	—
H16	1.00	1.00	1.00	1.00	1.00	1.00

Concerning the effects of variable disaggregation (*Run* B_t vs. C_t) for a given constant sample we generally find a constant or increasing trend in h_o^* scores for a given year. Apparently, this result occurs because output mix differences between aged and pediatric patients are explicitly considered in *Run* C_t, but not in *Run* B_t. When output mix differences are not taken into account (*Run* B_t), a DMU which produces, for instance, a high percentage of high cost, aged patient days may appear inefficient. When such differences are taken into account (*Run* C_t), much of the observed inefficiency can disappear (e.g., H9 in 1979) since the reference set DMUs and the DMU being evaluated are now more homogeneous with respect to output mix.

To summarize, many of the h_o^* values in Table 4 exhibit an increasing trend in a given year when moving from Runs A_t to B_t to C_t. Though the significance of this trend is difficult to assess, these findings suggest, at least tentatively, that proper variable specification and sample selection are important issues to be considered when using the DEA model in practical applications.[11]

SENSITIVITY ANALYSIS

Another important issue in the current study concerns the sensitivity of the DEA efficiency scores to variations in data inputs. These variations could result from DMU use of different accounting procedures to determine total routine costs, differences between the Medicare cost report and the AHA Annual Survey in reporting patient day statistics, etc. A difficulty emerges, however, in that data for DMU_{j_o} appear on both sides of the constraint equations. Furthermore, for efficient DMUs, the optimal basis will necessarily include DMU_{j_o}. Consequently, standard sensitivity analysis/parametric programming techniques available on most LP computer packages would appear useless in the present case.

However, as a partial solution to these problems, Charnes, Cooper, Lewin, Morey, and Rousseau [12] have defined the necessary conditions permitting the maximum possible decrease (increase) in a given output (input) for efficient DMU_{j_o} (holding all other inputs and outputs constant), while keeping DMU_{j_o} efficient without a change in its optimal basis. Observe that the sensitivity ranges provided by this method are conservative since a DMU may continue to be efficient even though a basis change occurs. These results are reported in Table 5 for the efficient DMUs discovered in *Run* A_t, $t=1,2$.

Overall, the sensitivity results with respect to outputs are inconsistent both across and within the DMUs, although for many DMUs, any decrease in certain outputs will cause a basis change (though not

Table 5: Sensitivity Analysis Results—Efficient DMUs in DEA *Run* A_t (Phase One)

	Aged and Pediatric Days			*Maternity Days*			*Other Routine Days*			*Total Inpatient Routine Costs*		
Efficient DMU	*Low* Possible*	*Decrease from Actual*	*% Decrease*	*Low* Possible*	*Decrease from Actual*	*% Decrease*	*Low* Possible*	*Decrease from Actual*	*% Decrease*	*High** Possible*	*Increase over Actual*	*% Increase*
H4 - 1978	7215	0	0	1043	0	0	2744	2373	46.4	$1,063,599	$ 88,077	9.0
H7 - 1978	13,417	3988	22.9	733	641	46.7	15,317	0	0	2,431,971	129,160	5.6
H15 - 1978	7656	0	0	840	0	0	14,809	0	0	2,223,494	240,685	12.1
H16 - 1978	11,693	0	0	2614	0	0	12,164	0	0	2,315,102	145,742	6.7
H1 - 1979	7156	0	0	611	0	0	0	4262	100.0	964,894	32,044	3.4
H4 - 1979	7400	0	0	905	0	0	3891	655	14.4	996,055	22,707	2.3
H6 - 1979	11,613	968	7.7	1405	0	0	14,694	0	0	2,206,935	47,095	2.2
H7 - 1979	14,844	1438	8.8	0	1169	100.0	13,477	0	0	2,353,449	108,560	4.8
H15 - 1979	8973	0	0	927	0	0	15,131	0	0	2,359,666	310,818	15.2
H16 - 1979	12,713	0	0	3107	0	0	12,126	0	0	2,301,608	254,579	12.4

*Lowest possible value for this output variable, holding all other outputs and inputs constant, and DMU_{j_o} remaining efficient without a basis change.

**Largest possible value for this input variable, holding all outputs constant, and DMU_{j_o} remaining efficient without a basis change.

necessarily a decrease in efficiency). Hence, it becomes difficult to draw any general conclusions concerning DEA's sensitivity to perturbations in outputs. Only if wide sensitivity ranges were found could we safely conclude that the efficiency ratings of the frontier DMUs were insensitive to variations in a particular output variable.

Likewise, concerning the single input variable, a relatively small variation ($< 10\%$) in total routine costs will cause a basis change to occur for several efficient DMUs. Thus, based solely upon the results in Table 5, it would be premature to conclude that the DEA efficiency ratings of efficient DMUs are sensitive or insensitive to changes in total routine costs.

On the other hand, for a given inefficient DMU, observe that each percentage variation in total routine cost translates into nearly a .01 change in its h_o^* score, holding all other factors constant.[12] That is, a 5% reduction in total routine costs for a given DMU, holding all else constant, will increase that DMU's h_o^* by almost .05. In light of the significant measurement error which exists in the determination of routine service costs (see [18]), at least a 5% variation in this variable for many hospitals would probably be expected. Consequently, the reasonableness and stability of the h_o^* scores and resultant cost savings estimates reported in Table 3 become suspect, and their interpretation must be tempered with the possibility that the large cost savings differences observed between the DEA and cost/patient day models are in part the result of measurement differences across hospitals.

PHASE TWO: COMPARISON OF 1978 AND 1979 EFFICIENT FRONTIERS

Applying the adjustments described in CCR [5], each hospital's inputs and outputs for Runs A_t-C_t, $t=1,2$ were adjusted to efficient levels as determined by each DMU's efficient frontier (1978 vs. 1979). DEA using Runs *A*–*C* was then performed on the combined 1978 and 1979 adjusted observations, and the resulting distribution of h_o^* scores examined for evidence of changes in efficient production possibilities over the 1978–1979 period. These results for *Run A* are displayed in Table 6 (in general, similar results were found for Runs *B* and *C*).

Inspection of Table 6 reveals that neither the 1978 nor 1979 time periods reflects greater potential for efficient operations.[13] Were the opportunities for efficient production of outputs greater in 1978 than 1979, we should expect a significantly higher percentage of 1978 DMUs to be rated efficient (i.e., $h_o^* = 1$). In this case, the percentages for both years are quite similar, and the average h_o^* score is .98 for both 1978 and 1979. The Table 6 results suggest that within the limitations of the variable sets

Table 6: Distribution of h_o^* Scores from *Run A* after Eliminating Managerial Inefficiencies (Phase Two)

Hospital	*1978*	*1979*
H1	.99	1.00
H2	.99	.97
H3	.93	.99
H4	1.00	1.00
H5	.99	.97
H6	.98	.99
H7	1.00	.96
H8	1.00	1.00
H9	.99	.99
H10	.97	.96
H11	1.00	1.00
H12	.93	.96
H13	.98	.96
H14	.99	.99
H15	1.00	1.00
H16	.91	1.00

examined, the overall potential for hospitals to achieve efficient operations did not differ significantly over the 1978–1979 period.

CONCLUDING REMARKS

The DEA methodology provides a model capable of aggregating the multiple inputs and outputs of nonprofit organizations into a single summary measure of efficiency. Importantly, the weights assigned to each input and output are not arbitrary, but based on Pareto Efficiency conditions. These developments represent significant advances in nonprofit performance measurement, since it is well-recognized that the absence of "profits" makes it difficult to measure the efficiency of nonprofit organizations.

Our research applies the DEA technique to efficiency evaluation of routine nursing services for a sample of Wisconsin hospitals. We compared the DEA results with those based on Medicare's efficiency criteria and cost/patient day measure.

Not unexpectedly, the findings differ under the two approaches. The efficiency measures themselves (h_o^* and cost/patient day) differ in that DEA accounts for differences in input requirements for DMUs with different output combinations, while the simple cost/patient day model

does not. Also, the routine cost limits defined by DEA are based on extremal relations and production function concepts—Medicare criteria are more arbitrary and designed to detect only grossly inefficient hospitals. Under DEA, the average potential routine cost savings per inefficient hospital was several times greater than that using Medicare criteria. Furthermore, over 60% of the DMUs in both 1978 and 1979 were inefficient per the DEA model, compared with none in 1978 and less than 7% in 1979 using Medicare criteria. Of the two models, DEA identified a much greater amount of inefficiency in the sample institutions.

One should not, however, jump to the conclusion that the large potential cost savings per the DEA model (Table 3) is entirely the result of inefficiencies which can be readily eliminated by hospital management. Certain limitations in our study prevent such a conclusion. First, case-mix differences across DMUs were not explicitly accounted for in the DEA model. DMUs with a similar patient day output mix may treat a significantly different mix of cases. Hence, a DMU's low h_o^* score may result from the DMU providing a more complex (and costly) case mix than that provided by efficient DMUs. A logical extension of this study is to re-examine the potential cost savings under DEA after case-mix output measures have been incorporated into the model.

Secondly, not all routine costs are controllable in the short run, so that resource reduction opportunities suggested by $(1-h_o^*)$ are overstated. Moreover, since hospitals might employ different combinations of fixed and variable inputs in producing given outputs, the actual short-run cost reduction which *is* possible via $(1-h_o^*)$ may differ across institutions.

In recognition of this problem, Medicare has recently begun excluding capital costs in setting routine reimbursement limits. Similar considerations could be incorporated into DEA by running the model with total routine costs split into controllable and uncontrollable elements, and then calculating the potential cost savings based only on those costs deemed controllable by hospital management.[14] Certainly, as found when outputs are disaggregated, consideration of disaggregated input measures would increase the h_o^* ratings of inefficient DMUs and reduce the cost-savings estimates calculated in Table 3.

Third, as noted previously, a great deal of measurement error exists in the total routine service costs variable. Since a hospital's efficiency is measured relative to other peer group institutions, the h_o^* scores and resulting costs savings estimates can be adversely affected by historical cost accounting, alternative accounting methods, and various arbitrary allocations used to derive total routine costs on the Medicare cost report. Moreover, since only a single aggregated input variable is considered, each percentage point change in total routine costs of an inefficient DMU

results in nearly a .01 change in its h_o^* score. Accordingly, the cost savings estimates given in Table 3 are also sensitive to routine cost variations, and should therefore be viewed with discretion. This finding suggests that DEA's usefulness in a hospital regulatory setting would be greatly improved by implementing price-level adjustments and uniform hospital accounting to reduce measurement error resulting from existing accounting practices. This conclusion is only tentative, however, since consideration of a more disaggregated input variable set could result in the insensitivity of certain h_o^* scores to variations in some of the disaggregated inputs.

Notwithstanding these limitations, the present study has demonstrated via a sample of Wisconsin hospitals, the fundmental differences between the DEA and cost/patient day approaches in weighting inputs and outputs and in discriminating efficient from inefficient institutions. Certainly, the methodological improvements discussed above would decrease the cost savings estimated in Table 3 for the DEA model. At the same time, however, the findings do suggest the DEA approach is both a practical and theory-based efficiency measurement model which deserves further empirical study in health care settings.

ACKNOWLEDGEMENTS

I wish to thank Werner Frank, Arie Lewin, Mark Covaleski, Don Hansen, Jim Mackey, John Surdick, Howard Thompson and Tom Williams for their helpful comments on earlier drafts of this paper. Also, I would like to thank Dave Dittman and Dick Morey for their computer assistance in completing the DEA runs.

NOTES

1. The reference here is primarily to output markets. Many nonprofits (e.g., governments, hospitals, universities, etc.) operate in fairly competitive input markets, while the output market, if it even exists, sets prices through some noncompetitive mechanism (e.g., rate regulation in hospitals). Hence, even if resources are purchased in competitive markets, the lack of market-determined output prices destroys the profit figure as a useful barometer of efficiency.
2. For instance, in 1976 the limitation was reduced from the 90th to the 80th percentile plus 10% of the median. More recently, for cost reporting periods beginning July 1, 1980, the limit was based upon 115% of the group mean and then further reduced to 108% of the group mean for periods beginning October 1, 1981. Other refinements have been introduced (e.g., see [9]), yet the rather arbitrary nature of the limit remains.
3. See [10] p. 686 for further elaboration of this point.

4. As shown in [5], a DMU is Pareto Efficient only if $h_o^* = 1$ *and* all slacks from (4.1) and (4.2) are zero. Hence, in reality, input reduction via h_o^* only, is not always sufficient to position an inefficient DMU on the efficient production surface. However, as indicated in [4], the slack amounts represent "corners of a frontier" so that input reduction via h_o^* will make a DMU nearly efficient (i.e., h_o^* almost equal to one). Thus, for purposes of comparison with the input reduction possible under the Medicare model, it would seem little is lost by ignoring the slack values resulting from (4.1) and (4.2).
5. Readers interested in further mathematical and economic concepts underlying the model should consult [5]. Additional discussion and interpretation of model outputs not considered here are also provided in [4,10,11].
6. Example classification models are (1) the AID system [13], (2) the AHA Clustering Model [14], and (3) the Washington State Clustering Model [16, 17]. Critical reviews of various hospital classification models are provided in [15,17,18].
7. Much criticism has been directed at Medicare's classification methodology for being overly simplistic and crude (e.g., see [17,18]). Thus, to obtain hospitals more homogeneous in output than those resulting from the Medicare groupings, only clusters generated by the Wisconsin classification model were considered. The clustering variables used in the Wisconsin model are as follows: (1) inpatient volume for basic services (e.g., adult med./surg., intensive care, burn care, etc.): (2) intensity of basic inpatient services; (3) special services (e.g., respiratory therapy, organ bank, open heart surgery, etc); (4) outpatient equivalent days; (5) educational programs (e.g., residents, nursing school, allied education); (6) government days (e.g., Medicare and Medicaid utilization). For additional information concerning variable measurement and the clustering method used, see [19].
8. Effective October 1, 1981, the nursing salary differential was reduced to 5%.
9. Having made this observation, we might not question the appropriateness of ranking all DMUs by h_o^* as done in Table 2. Two hospitals, each producing a significantly different output mix, may have different opportunities for resource reduction and/or output augmentation. An alternative then is to examine efficiency rankings for DMUs having the same reference set (i.e., optimal basis) of efficient DMUs. This procedure was not performed however, since the small sample size yielded only a few DMUs with the same optimal basis.

   We should point out though, that the h_o^* rankings provided in Table 2 could be argued as still meaningful since operationally $(1-h_o^*)$ measures the % input reduction required for DMU_{j_o} to produce its given output as efficiently as the DMUs in its reference set. Geometrically, $(1-h_o^*)$ indicates the relative distance DMU_{j_o} must travel along its own contraction path to be positioned on the efficient production surface. Hence if h_o^* for $DMU_x > h_o^*$ for DMU_y, we know that DMU_y must contract its resources proportionately more than DMU_x in order to be considered Pareto Efficient with respect to its individual reference set.
10. Large differences in potential cost savings persist even when the Medicare criteria are updated to reflect recent regulation changes. For instance, when the cost/patient day limit is set at 108% of the group mean (see note 2), 4 DMUs (25%) in each year are considered inefficient with a total cost savings of $813,992 in 1978 and $1,027,858 in 1979, still quite a bit less than that determined for DEA *Run* A_t.

11. Readers are referred to Lewin, Morey, and Cook [20] who present a general methodology for selecting variables to be considered in DEA efficiency evaluation.
12. An inefficient DMU's h_o^* score does not change *exactly* .01 for each percentage point change in total routine costs. Specifically, for a 1% change in total routine costs, the new h_o^* equals the original $h_o^* \div .99$. For instance, for a DMU with original h_o^* of .90 a 1% decrease in total routine costs will result in a new h_o^* score of $.90 \div .99 = .909$.
13. Other modes of comparison such as Kullback-Leibler divergence statistic (see [21]) may also be useful in comparing distributions of h_o^* scores. For purposes of the present study, simple inspection of the Table 6 results would seem to provide an adequate basis for our conclusions.
14. This procedure was not performed for two reasons. First, to do so would require the nontrivial task of identifying and excluding all uncontrollable costs from total hospital costs and then redoing the overhead allocation for each DMU in the sample. Second, for the time period examined (1978–1979), Medicare did include capital costs in calculating routine limits. Hence, for purposes of comparisons, capital costs were not eliminated in performing the DEA runs.

REFERENCES

1. *Objectives of Reporting by Nonbusiness Organizations.* 1980. Stamford, Conn: Financial Accounting Standards Board, FASB Financial Accounting Concepts Statement No. 4.
2. Hadley J. Research in Health Manpower Productivity: A General Overview. In John Rafferty (ed). *Health Manpower and Productivity.* 1974. Lexington MA: Lexington Books, 147–150.
3. Anthony RN, R Herzlinger. *Management Control in Nonprofit Organizations.* 1975. Homewood IL: Richard Irwin, Inc.
4. Lewin AY, RC Morey. Measuring the relative efficiency and output potential of public service organizations: An application of Data Envelopment Analysis, *International Journal of Policy Analysis and Information Systems* (1981) 5:(4)267.
5. Charnes A, WW Cooper, E Rhodes. Measuring the efficiency of decision making units, *European Journal of Operational Research* (1978) 2:(6)429. See also "Corrections," *op. cit.* (1979) 3:(4)339.
6. Berki S. *Hospital Economics.* 1972. Lexington, MA: Lexington Books.
7. Public Law No. 92-603, § 223, Social Security Amendments of 1972, 92nd United States Congress.
8. *Medicare Provider Reimbursement Manual, Part I* (HIM-15), U.S. Department of Health, Education and Welfare, Social Security Administration.
9. Bowman RW. Proposed medicare regs would reduce hospitals' reimbursement. *Modern Healthcare* (July 1979) 9:(7)75–76.
10. Charnes A,WW Cooper, E Rhodes. Evaluating program and managerial efficiency: An application of Data Envelopment Analysis to program follow through, *Management Science* (June 1981) 27:(6)668.
11. Bessent AM, EW Bessent. Determining the comparative efficiency of schools

through Data Envelopment Analysis, *Educational Administration Quarterly* (Spring 1980) 16:(2)57.
12. Charnes A et al. Data Envelopment Analysis: A non-Archimedean proof and robustness calculations, Working Paper, Fuqua School of Business, Duke University, May 1981.
13. Neumann BR. An empirical investigation of the relationship between an AID hospital classification model and accounting measures of performance, *Journal of Accounting Research* (Spring 1979) 17:(1)123.
14. Joseph P, RN Iyer. Classification of community hospitals, *Health Services Research* (Winter 1975) 10:(4)349.
15. Neumann BR. Commentary—On the use of hospital classification models," *Inquiry* (Winter 1979) 16:(4)356.
16. Klastorin T, C Watts, V Trivedi. *A Study of The Classification of Hospitals for Prospective Reimbursement.* January 1978. Report No. 10, Health Care Financing Research and Demonstration Series, Health Care Financing Administration.
17. Trivedi VM. Classification of short-term general hospitals for control and equity, *Inquiry* (September 1978) 15:255.
18. Ruchlin HS. Problems in measuring institutional productivity, *Topics in Health Care Financing* (Winter 1977) 4:(2)13.
19. Wisconsin Hospital Rate Review Standards Development Committee. Hospital grouping—Background and methodology, 1979, State of Wisconsin.
20. Lewin AY, RC Morey, TJ Cook. Evaluating the administrative efficiency of courts, *Omega* (1982) 10:(4)401.
21. Charnes A, WW Cooper. Management science relations for evaluation and management accountability, *Journal of Enterprise Management* (1980) 2: (2)143.