UNIVERSITY OF OSLO
HEALTH ECONOMICS
RESEARCH PROGRAMME

Testing DEA Models of Efficiency in Norwegian Psychiatric Outpatient Clinics

Sverre A.C. Kittelsen
Frisch Centre

Jon Magnussen
SINTEF Unimed NIS
Health Services Research

Working Paper 1999: 4

SINTEF Unimed NIS
Health Services Research

Testing DEA Models of Efficiency in Norwegian Psychiatric Outpatient Clinics

Sverre A.C. Kittelsen, Ph. D.
Frisch Centre, and HERO – Health economic research programme at the University of Oslo

Jon Magnussen, Ph.D.
SINTEF Unimed NIS Health Services Research, Trondheim

Health Economics Research programme at the University of Oslo
HERO 1999

Paper presented at the 6th European Workshop on Efficiency and Productivity Analysis
Copenhagen, Denmark, October 29-31, 1999

Health Economics Research programme at the University of Oslo 1998- 2003
ISSN 1501-9071, ISBN 82-7756-036-2

Abstract

While measures of output in mental health care are even harder to find than in other health care activities, some indicators are available. In modelling productive efficiency the problem is to select the output variables that best reflect the use of resources, in the sense that these variables have a significant impact on measures of efficiency.

The paper analyses cross-sectional data on the psychiatric outpatient clinics of Norway using the Data Envelopment Analysis (DEA) non-parametric efficiency measurement method, and tests the variable specification using statistical tools recently introduced in the literature. In addition to outputs, the importance of different profession or educational groups on efficiency is examined, and results are compared for separate samples of clinics for children and youths (BUP) with clinics for adults (VP).

1. Introduction

In the past 5-10 years there has been major structural changes in the provision of mental health care services in Norway. A central element in this restructuring has been a shift of focus from inpatient to outpatient care. Thus in the 5 year period 1992-96 the number of beds in psychiatric nursing homes was reduced by 36 percent, the number of hospital beds reduced by 4 percent and the number of outpatient visits increased by 25 percent (Hagen, 1998). The important thing to note about this development, is that it is *intended*, and not simply due to a change in the demand for services.

The main purpose of the restructuring can be described as:

- *Decentralisation* of services. Rather than providing services in a few centrally located institutions, people should have access to care as close to home as possible.
- *Differentiation* of services. Rather than only providing psychiatric care in hospitals or nursing homes, care is now given in variety of settings: Outpatient clinics, district psychiatric centers, small apartment groups with adjoining treatment facilities, as well as the more traditional hospitals and nursing homes.
- *Accessibility* to services. An important motivation for the decentralisation is that accessibility to services should not depend on geographical location.

In principle the aim of this is to provide a better set of services to patients in need of psychiatric services. The underlying motivation is also one relevant for a cost-effect analysis: It is assumed that the same set of patients can be provided for more efficiently in a decentralised, outpatient based system, than in the "old" inpatient based system. This notion is to some degree supported by evidence. Goodwin (1997) sites several studies that find community based care to be less costly than institutional care.

Even though there has been a substantial increase in the capacity of outpatient clinics, it is still a stated political goal to increase outpatient activity. The demand for outpatient services greatly exceeds what seems to be the sectors present capacity. Thus questions have been raised as for whether excess demand could be met not only by increased capacity but also by a more efficient use of resources. Concern about low levels of efficiency have been fuelled by simple measures of labour productivity: In outpatient clinics for adults average number of consultations per therapist day is 1,71 in clinics for children and young, the average number is 1.12.

While these numbers seem low, both relevant measures of output and the relationship between inputs and outputs is difficult to establish for production of mental health services. Thus it may be argued that seemingly low labour productivity rates simply reflect that we only measure one dimension of the health care activities of the outpatient clinics, or else that the treatment process in itself is so complex that the size of these numbers is to be expected.

Allthough we provide some evidence, the purpose of this paper is not to resolve these issues. Instead we seek to illuminate two policy questions related to the production of care in outpatient clinics. Firstly the question of the existence of scale economics may affect the dimensioning and therefore the geographical structure of the supply of outpatient services. Secondly the relationship between staffing mix and productivity may help to explain observed differences and to draw implications about optimal staffing mix.

There is a clear distinction between outpatient services for children and young people and outpatient services for adults. These services are also delivered in different settings; outpatient clinics for children and youth (hereafter termed BUP) and outpatient clinics for adults (hereafter termed VP). In this paper we analyse cross-sectional differences in efficiency using the Data Envelopment Analysis (DEA) non-parametric efficiency measurement methods. Furthermore we test the variable specification using statistical tools recently introduced in the literature. The remainder of the paper is structured as follows. In section 2 we briefly describe the characteristics of the two outpatient settings, and the definitions and measurement of inputs and outputs. Section 3 describes the efficiency concepts, the hypotheses to be tested and the statistical tools used in the tests. In section 4 the results of the DEA analysis are presented. Section 5, finally, provides a discussion and some concluding comments.

2. Outpatient care for mental health patients.

2.1 The treatment process

The distinction between psychiatric services for children and youths and psychiatric services for adults reflects both on the type of treatment and on the setting in which treatment is provided. Loosely formulated psychiatric care for children and youths is aimed at correcting an undesired behavioural pattern through the combined use of therapy and interaction with the patient’s environment (relatives, school etc). In Norway more than 95% of those children and youths who receive psychiatric care do so in an outpatient setting. Psychiatric care for adults is, on the other hand, delivered both in institutions and in an outpatient setting. Adult patients are also often made subject to medication (which is nearly absent in care for children and youths), and a higher proportion of these are patients with chronic illnesses. Outpatient services are therefore delivered in separate settings; clinics for children and youths (BUP) and clinics for adults (VP).

The treatment process will consist of a series of interventions related to each patient. The interventions will be of different forms depending on the type of disorder, the social setting and also as we shall later argue on the outpatient clinic itself. It is important to be aware that mental disorders, in contrast to most somatic disorders, are difficult to diagnose. Also there is very little consensus as to how specific disorders should be treated; leading to a situation where professional and cultural environment in some cases may be a better predictor of type of treatment that the diagnose itself. This is likely to affect productive performance, a point to which we will return later.

While the differences between them implies that BUP and VP have to be treated separately, they are largely staffed by the same professions, and we can also provide a discussion of defining and measuring outputs and inputs in a general setting.

As noted the treatment process will consist of a series of interventions. These interventions will be performed by different types of staff and be directed at different goals. Interventions may be aimed directly at the patient or also at the patient's surroundings (schools, relatives, primary health care etc). They can furthermore be done in a situation where the patient and therapist is alone, or in various forms of group settings.

2.2 Measuring inputs and outputs

Ideally one would like to model the input-output relationship using data on number of interventions by type and number of FTEs by category. While FTEs are available on a fairly detailed level, number of interventions are not. Furthermore there are differences between BUP and VP in the quality of data that are available. In this paper the following output definitions are chosen.

Children and youths (BUP):
Output data for the BUP sector are fairly good. While data on the types of intervention is missing, we have data on the total number of interventions, total number of patients and total number of therapist hours used on patient related activities. We therefore define three types of outputs:

Y_T – interventions
Y_P – patients
Y_H – hours

Depending of the type of problem, each patient will receive a number of interventions, each intervention implying a certain number of therapist hours. Including all three outputs allows us to compare efficiency in clinics where a small number of patients receive a large number of interventions with clinics where a large number of patients receive a small number of interventions. Note, however, that we do assume that there are no inefficiencies in the chosen treatment programmes. Thus every hour of every intervention is assumed to be "necessary" and equally valuable to the patient.

Adults (VP):
For the outpatient clinics data are not available forpatients, interventions nor hours. We do, however, have information about the total number of consultations, i.e. patient visits. Thus we model only one output for VP:

Y – number of patient visits

We use similar input definitions for both BUP and VP. Thus we define:

X_U – university educated staff

X_H – college educated staff

X_R – other staff

The latter variable is only available in the VP sector. College education in Norway includes the three or four year educations of nurses, social workers etc, while psychologists, psychiatrists and physicians are university trained. Thus for BUP our initial model consists of three outputs and two inputs, while our VP model consists of one output and three inputs.

3. Estimating efficiency

The idea of measuring technical efficiency by a radial measure representing the proportional input reduction possible for an observed unit while staying in the production possibility set stems from Debreu (1951) and Farrell (1957) and has been extended in a series of papers by Färe, Lovell and others[1]. Farrell's specification of the production possibility set as a piecewise linear frontier has also been followed up using linear programming (LP) methods by Charnes, Cooper et al[2]. The decomposition of Farrell's original measure relative to a constant returns to scale (CRS) technology into separate measures of scale efficiency and technical efficiency relative to a variable returns to scale (VRS) technology is due to Førsund & Hjalmarsson (1974) and has been implemented for a piecewise linear technology by Banker, Charnes and Cooper (1984). Their DEA formulation has served as the main model of most recent efficiency studies and is the basic model in this paper.

Various measures of productive efficiency are possible, such as social efficiency and allocative and cost efficiency, which we are not able to estimate due to lack of data on prices and/or social evaluation of production. Instead we concentrate on technical measures of efficiency, in the sense that we compare actual behaviour with some point on the frontier of the technically feasible set. This frontier point will in general not be the optimal behaviour if values are applied, but if the model is correctly specified.the optimal behaviour will be one of the points on the frontier.

[1] E.g. Färe & Lovell (1978) and Färe, Grosskopf & Lovell (1985).

Technical efficiency can be measured both in an input direction, as the proportion of inputs that are nescessary to produce a given level of output, and in an output direction, as the ratio of actual production to the maximum production given the level of inputs. In the psychiatric outpatient clinics we have chosen to concentrate on the latter, implying a focus on how much more psychiatric treatment could be provided with existing levels of staffing, if clinics were technically efficient. Using the terminology of Førsund & Hjalmarsson (1974), this Farrell (1957) radial estimate of technical output efficiency is defined by

$$\hat{E}_2 = \text{Min}_\theta \left\{ \theta \,\middle|\, (\mathbf{y}/\theta , \mathbf{x}) \in \hat{P} \right\}, \qquad (1)$$

where $\mathbf{y}$ is a vector of K outputs and $\mathbf{x}$ is a vector of L inputs, and $\hat{P}$ is an estimate of the production possibility set or technology

$$P = \left\{ (\mathbf{y},\mathbf{x}) \in \Re_+^{K+L} \,\middle|\, \mathbf{y} \text{ can be produced from } \mathbf{x} \right\}, \qquad (2)$$

[2] E.g. Charnes, Cooper & Rhodes (1978) who originated the name DEA. For an overview of the literature on DEA see e.g. Seiford (1996).

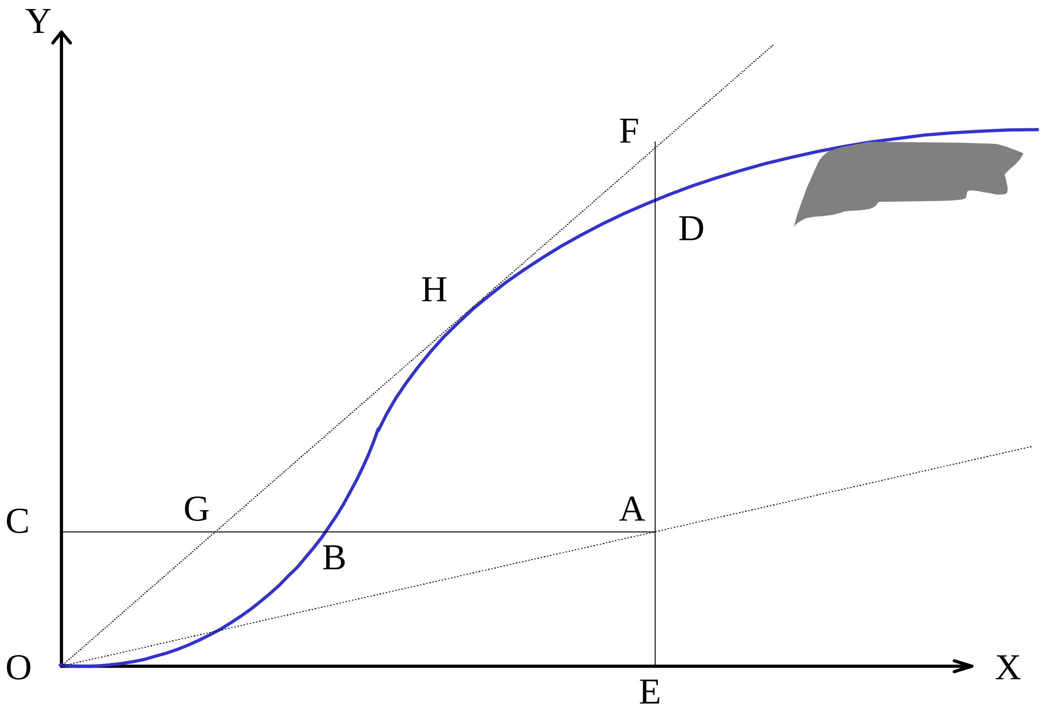


Figure 1: Efficiency measures in input-output space. E_2=EA/ED, E_3=EA/EF, E_5=ED/EF.

Figure 1 illustrates the basic concepts. Point A is an observed input/output combination in a one-input one-output technology, and the technology set is the area below and to the right of the curved frontier. Given a constant level of input OE , the technical output efficiency of unit A is the ratio of actual output EA (=OC) to the maximum production that is feasible ED.

The figure also illustrates the measure of technical productivity E_3 that is the ratio of the output-input ratio of observation A, the slope of the dashed line OA, and the maximal output-input ratio, the slope of the dashed line OH. Geometrically this can be seen to be equal to the ratio EA/EF. Technical productivity is sometimes termed gross scale efficiency, implying a comparison of actual production per unit of input behaviour to the maximal production per unit of input had the production taken place at the technically optimal scale of point H. The estimate of this measure can be formulated

$$\hat{E}_3 = \text{Min}_\theta \left\{ \theta \,\middle|\, \left({\gamma \mathbf{y}}/{\theta}, \gamma \mathbf{x} \right) \in \hat{P} \right\}, \qquad (3)$$

where γ is a free scalar. The inverse of the optimal value of γ is the scale indicator that measures the proportion of actual inputs to the inputs at the optimal scale (i.e. OF/OH in figure 1). Finally we introduce the pure scale efficiency measure E_5 which is the ratio of the productivity of the technicall efficient frontier point and the maximal productivity (i.e. EA/EF in figure 1). The estimate is defined simply by

$$\hat{E}_5 = {\hat{E}_3}/{\hat{E}_2} \qquad (4)$$

One may note that if the production technology exhibits constant returns to scale (CRS), the frontier is a straight line from the origin, and the measures of technical efficiency and technical productivity coincide ($E_2=E_3$). This also implies that all observations are scale efficient ($E_5=1$).

The DEA estimate of the production possibility set is given by a set of linear constraints

$$\hat{P} = \left\{ \mathbf{Y}\lambda \geq y, \mathbf{x} \geq \mathbf{X}\lambda, \lambda \geq 0, \sum_{i \in N} \lambda_i = 1 \right\}, \qquad (5)$$

where $\mathbf{Y}, \mathbf{X}$ are the vectors or matrices of observed outputs and inputs and λ is a vector of reference weights. This corresponds to the formulation in Banker, Charnes & Cooper (1984), and is the minimum extrapolation estimator of the technology satisfying convexity, free disposability of inputs and outputs and feasibility of observed units (Banker, 1993), as illustrated in figure 2.

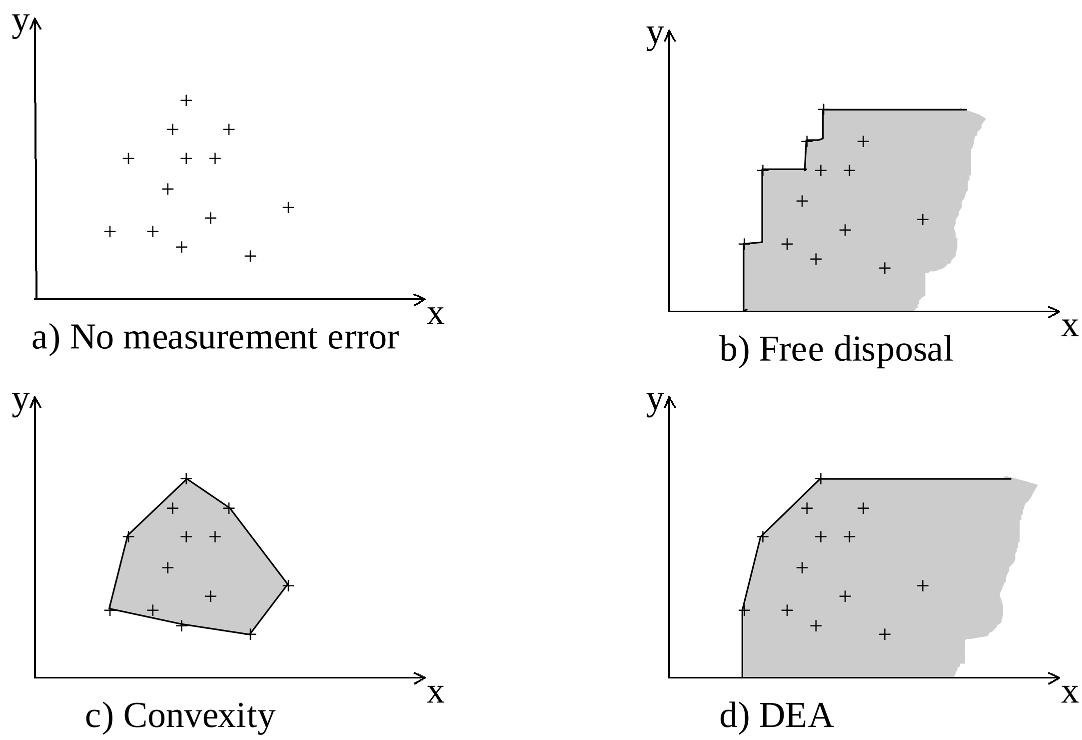


Figure 2: The DEA assumptions on the possibility set

4. Testing DEA models

Statistical tests have been few in the DEA literature. Valdmanis (1992) among others has used the Mann-Whitney rank-order test to compare the efficiency of public vs. not-for-profit hospitals and find the public hospitals significantly more technically efficient in seven out of ten different input-output specifications. While her approach is fruitful in assessing the performance of separate groups and demonstrates the robustness of results across specifications, her method does not give an answer to the question of which specification is best.

Farrell (1957) recognised that statistical tests should be based on the frequency distribution of efficiencies. The problem is that when one assumes that all observations are feasible, i.e. no measurement error, any sampling error would bias the DEA efficiency estimators upwards, since the true frontier in general lies outside the estimated frontier. However, recognising that sampling error exists in DEA analysis, also gives a basis for statistical analysis of ”deterministic” frontiers.

While tests such as the Man-Whitney rank-order tests have been used for subset comparisons[3], the assumptions underlying most tests are not fulfilled when testing model specification since such models generally will be nested. A model 0 will be nested within another model 1 if model 0 can be obtained from model 1 as a special case. This implies that a CRS model is nested within a VRS model, an aggregated model is nested within a dissagregated model, and a model without a specific variable is nested within a model wich includes this variable. In nested models, the DEA estimates of efficiency will be ranked so that $\hat{E}^1 \geq \hat{E}^0$ for every observed unit, implying that the bias of the estimators will be at least as large for model 1 as for model 2, and usually larger. Any simple test based on the difference or ratio of such estimators will therefore also be distorted.

In recent developments, Banker (1993) has proven the consistency of the DEA estimators under specific assumptions and suggested statistical tests of model specification, while Korostelev, Simar and Tsybakov (1995a, 1995b) have been concerned with the rate of convergence of non-parametric frontier estimators. Kneip, Park and Simar (1996) extends these results to a more general model. Simar and Wilson (1995) suggests a bootstrap method for estimating the bias and confidence intervals of efficiency estimates and Simar and Wilson (1997) extend this to suggest a test of returns to scale[4]. Even though this approach seems feasible, it would be advantageous if simpler techniques were available.

So far, no tests have been suggested that can be shown analytically to able to discriminate between competing models, especially in small samples. While suggesting among others the Kolmogorov-Smirnov test used below, Banker (1993, p.1272) warns that “... the results should be interpreted very cautiously, at least until systematic evidence is obtained from Monte Carlo experimentation with finite samples of varying sizes”. Banker (1996) has summarised a series of Monte Carlo runs, using 10-30 repetitions in each evaluation, while Kittelsen (1999) has extended this to 1000 repetitions. The results indicate that some tests give crude but usable approximations of the true significance level and power functions, except in very small samples. Of the tests evaluated, the Kolmogorov-Smirnov test is the most conservative, while the ordinary T-test of the difference of means has more power, but tends to more easily overreject a true null hypothesis in small samples and high dimensionality. Banker (1993) also suggested two F-tests which yield similar results to the Kolmogorov-Smirnov test, but unlike the latter, these F-tests are based on specific assumptions on the distribution of inefficiency, and are not reported here.

If no parametric assumptions are maintained about the inefficiency distributions, the Kolmogorov-Smirnov nonparametric test of the equality of two distributions is a suitable approximation. Applied to the distributions of i.i.d. efficiency estimates, and denoting the estimated cumulative distribution function of these as $S^0(E), S^1(E)$, the statistic

$$D^+ = \mathrm{Max}_E \left\{ S^0(E) - S^1(E) \right\} \tag{6}$$

---

[3] See e.g. Valdmanis (1992) or Magnussen (1996).
[4] See Grosskopf (1996) for a survey of statistical inference in nonparametric models.

is asymptotically distributed with a rejection probability of

$$\Pr\left[D^{+} > \left(\frac{n^0 n^1}{n^0 + n^1} \right)^{-\frac{1}{2}} z \right] = e^{-2z^2}, \quad z > 0 \tag{7}$$

which makes it applicable for testing one-sided hypotheses (Johnson & Kotz, 1970).

For comparison, the simple T-statistic[5] for the equality of group means is reported:

$$T = \frac{\text{Mean}_{i \in N^1}\left(\hat{E}_i^1\right) - \text{Mean}_{i \in N^0}\left(\hat{E}_i^0\right)}{\sqrt{\frac{n^1 \,\text{Var}_{i \in N^1}(\hat{E}_i^1) + n^0 \,\text{Var}_{i \in N^0}(\hat{E}_i^0)}{n^1 + n^0 - 2}\left(\frac{1}{n^1} + \frac{1}{n^0}\right)}} \tag{8}$$

which, if sample means are i.i.d. normal, is T-distributed with $n^0 + n^1 - 2$ degrees of freedom. By the central limit theorem the sample means will be approximately normal unless sample size is very small. The expression greatly simplifies when $n^0 = n^1$ as is the case in the reported simulations.

[5] See e.g. Bhattacharyya & Johnson (1977, p.295-296).

5. Data and results

Data are collected from 79 VPs and 49 BUPs, and taken from SAMDATA (1998). Tables 1 and 2 presents some summary statistics.

Table 1 : Summary statistics for the sample of 51 BUP clinics.

	Average	StDev	Min	Max
Outputs				
Y_T Consultations	3 034	2 368	463	11 863
Y_P Patients	302	206	44	1 047
Y_H Hours	5 603	5 087	1 053	27 911
Inputs				
X_U University educated staff	6.1	5.0	-	25.0
X_H College educated staff	5.7	5.6	1.8	39.4
Derived data				
X_{UH} Therapists	11.8	10.1	2.0	64.4
Y_T/Y_P Consultations per patient	10.1	3.2	5.0	21.5
Y_H/Y_T Hours per consultation	1.77	0.31	0.99	2.63
Y_H/X_{UH} Hours per therapist	473.6	152.2	232.2	877.0
$Y_T/(X*230)$ Consultations per therapist day	1.18	0.38	0.47	2.00
X_U/X_{UH} University share of staff	0.50	0.15	-	0.75

Table 2: Summary statistics for the sample of 79 VP clinics.

	Average	St.dev.	Min	Max
Output				
Y Number of consultations	5 447	4 391	1 077	27 177
Inputs				
X_U University educated staff	7.7	5.7	1.5	34.7
X_H College educated staff	6.2	5.2	-	34.4
X_R Other staff	2.8	2.4	-	14.9
Derived data				
X Total staff	16.7	12.1	5.0	84.0

X_{UH} Therapists	13.9	10.1	4.0	69.1
$Y/(230*X_{UH})$ Consultations per therapist day	1.71	0.55	0.75	3.63
X_U/X University share of staff	0.46	0.11	0.25	0.78

On the whole, the VP clinics tend to be slghtly larger than the BUP clinics, with more staff and more consultations. The number of consultations per therapist is also larger in the VP clinics, but the striking feature is the large productivity variability in both sectors. Dividing the average number of consultations per therapist with the maximum, a partial measure of mean therapist productivity is 59% in the BUP clinics and 47% in the VP clinics. The question at hand is whether the available data lets us specify a model richer in scale properties and input-output specification and so help explain some of this variability, and give us better estimates of efficiency and productivity.

The procedure we follow is to first specify a model with variable returns to scale and with all available variables included. Next we exclude one variable at a time, and test whether the variable has a significant impact on the estimated efficiencies. The null hypothesis is in each case the conservative choice that the variable has no significant impact. If the test statistic is less than the critical value, the null yhypothesis is accepted, and the variable in question is excluded from the model. A similar procedure is used for testing for aggregation, where allowing aggregation is the null hypothesis, and for testing returns to scale, where CRS is the null hypothesis. Since the sample size is fairly small, and the T-test tends to overreject in small samples, we choose the Kolmogorov-Smirnov statistic D^+ as the descisive statistic. On the other hand we do not want to accept the null too easily, so we will use a 5% rejection level.

Table 3: Tests for BUP model

	Change in mean E_2	KS-test	T-test	Conclusion
Critical values 5%		0.242	1.660	
Critical values 1%		0.300	2.364	
H_0: Exclude patients	0.022	0.118	0.521	Accept
H_0: Exclude hours given patients excluded	0.021	0.098	0.491	Accept
H_0: Exclude both patients and hours	0.043	0.176	1.021	Accept
H_0: Constant returns to scale (CRS)	0.050	0.157	1.202	Accept
Result: (Y_T,X_U,X_H,CRS)				

The results of the model specification tests are reported in tables 3 and 4. In the BUP model of table 3, the initial full model is (Y_T,Y_P,Y_H,X_U,X_H,VRS). Both the number of patients and the number of hours are excluded from the model since they do not have a significant impact on efficiencies. The number of consultations Y_T is a priory taken to be the main output, and its inclusion is not tested since the model would not have an output without this variable. The results do not in itself mean that the number of patients or hours are not costly in terms of resource usage, but may rather indicate that they are sufficiently correlated with the number of consultations so that this latter variable is a sufficient indicator of the level of production in the BUP clinics.

Similarly, CRS is not rejected as a scale assumption, indicating that the size of a clinic does not significantly affect productivity. The resulting model is the most conservative, since all null hypotheses were accepted. Even though a sample size of 51 makes rejection difficult in any statistical analysis, the result seems reasonably robust since the statistic values are quite far from the critical levels.

Table 4: Tests for VP model

	Change in mean E_2	KS-test	T-test	Conclusion
Critical values 5%		0.195	1.655	
Critical values 1%		0.241	2.350	
H_0: Aggregate X_{UH}=X_U+X_H	0.065	0.203	1.840	Reject
H_0: Aggregate X_{HR}=X_H+X_R	0.057	0.177	1.632	Accept
H_0: Aggregate X=X_U+X_{HR}	0.080	0.215	2.435	Reject
H_0: Constant returns to scale (CRS)	0.109	0.253	3.443	Reject
Result: (Y,X_U,X_{HR},VRS)				

For the VP model we only have data for one output, and the analysis is instead concentrated on permissable aggregation of inputs. From economic theory, no input should be excluded, but one may instead ask whether the inputs are sufficiently homogenous as to be summed into fewer variables. The inital full model is (Y, X_U, X_H, X_R,VRS), and each pair of adjacent labour inputs is summed. The aggregation of the therapists with college and university education is rejected, allthough not strongly. The alternative aggregation of college educated staff and other staff is accepted, but a full summation of all staff categories is not. Finally a CRS specification is rejected quite strongly. In the resulting accepted model, the only simplification was the summation of college educated and other staff.

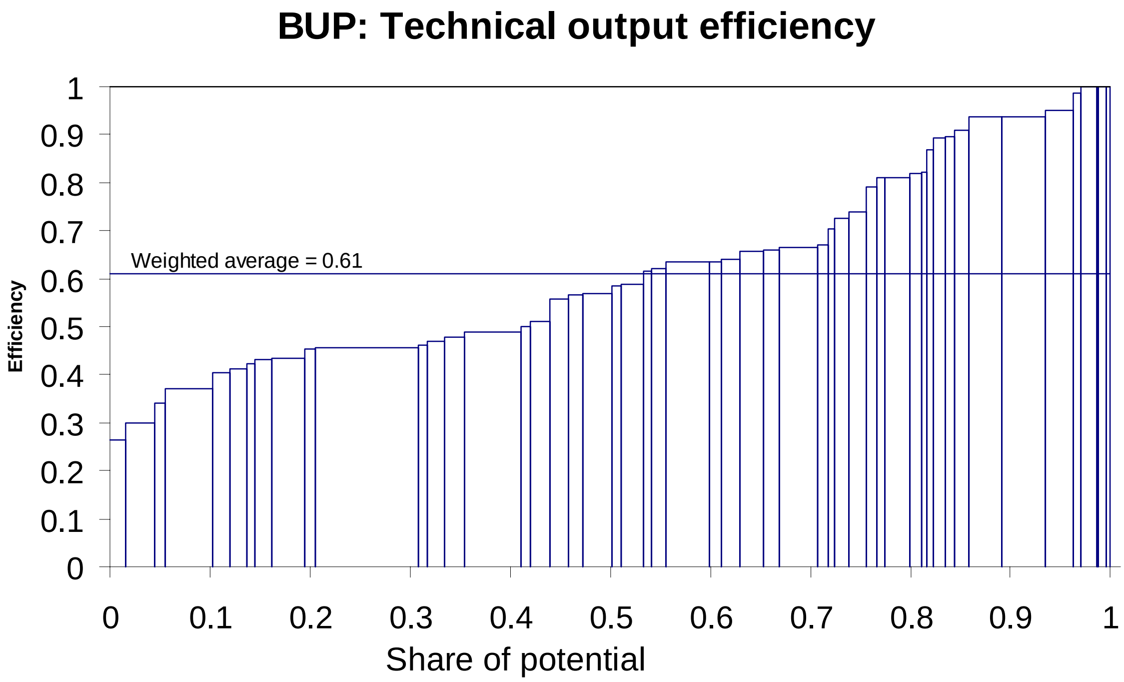
Table 5: Main results for BUP model

	Average	Stdev	Min	Max	Structural	Industry
$E_2=E_3=E_5$	0.656	0.209	0.264	1.000	0.575	0.611
MP(X_U)=$\partial Y_T/\partial X_U$	465	238	-	906	606	445
MP(X_H)=$\partial Y_T/\partial X_H$	494	367	231	1 231	275	518

The main efficiency results and other properties of the estimated technology are given in tables 5 and 6. Since the accepted model for the BUP clinics was CRS, the efficiency, productivity and scale measures collapse into a single estimate for each clinic. The average of estimated clinic efficiencies is 66%, but the variability is still large. In addition to the mean and spread of clinic efficiencies, the ”structural” measures is defined by the estimates for the average clinic, while the ”industry” measures is the output weighted mean. Both these efficiencies are lower than the unweighted average. Finally the marginal product of each labour input on the frontier mapping of each clinic is reported, revealing an estimate of how many more consultations an efficient clinic could undertake if it increased its staffing in that category by one position. Interestingly, this is on average greater for college educated than for university personnel. Because of the piecewise linear structure of the DEA estimate of the frontier, one should excecise care in the interpretations of MPLs for individual clinics, but average results are still of interest.

Table 6: Main results for VP model

	Average	Stdev	Min	Max	Structural	Industry
E_2	0.650	0.212	0.274	1.000	0.581	0.656
E_3	0.542	0.182	0.247	1.000	0.529	0.537
E_5	0.844	0.121	0.482	1.000	0.911	0.820
Scale indicator $1/\gamma$	2.063	1.567	0.375	9.395	2.125	2.944
MP(X_U)=$\partial Y_T/\partial X_U$	675	601	-	2 698	410	517
MP(X_{HR})=$\partial Y_T/\partial X_{HR}$	600	492	-	4 030	600	531



The mean technical output efficiency for the VP clinics is quite similar to that of the BUP clinics, but in this case the weighted industry efficiency is about the same as the unweighted mean, implying that large and small clinics are on average equally inefficient. From a productivity point of view, however, the smaller clinics fare better. Since CRS was rejected in this sector, the mean technical productivity of 54% is smaller than the mean efficiency estimate, with a pure scale efficiency of 84%. The scale indicator is on average considerably greater than 1, implying that most clinics are larger than the optimal size. While the mean marginal product of university educated staff is greater than that of other staff, both have great variability.

Figure 3: Distribution of efficiency and productivity estimates in the BUP clinics

6. Discussion

The main results emerging from this analysis are:

- *Industry efficiency is around 60% in the BUP sector and 66 % in the VOP sector.*

Based on these results there seem to be considerable room for improved activity in these clinics. It is also interesting, although probably coincidental, that the potential for increased output is not that far from the officially stipulated goal of 50% increased productivity (St.meld 25, 1996-97). As for the BUP sector it should also be remembered that these measures are derived under the assumption that medical practice is efficient. If this is not the case the observed best-practice and the theoretical frontier will not coincide, and there is room for further improvement in outputs.

There are, however, some qualifying remarks that need to be made. Firstly there may be variations between clinics to what extent time should be spent on treating outpatients. In some cases personnel are dedicated to other tasks such as "barnevern" or inpatient activities at adjacent hospitals. Also there will be variations in to what extent personnel at clinics spend their time servicing the primary health care. We do not capture "consultative work" as an output in our model, and neither are we able to correct the input measures for time spent in other facilities. The implication of this is both that the clinics on the frontier may not be the "real" reference units, and also that the potential for output improvement of inefficient clinics is less that what emerges from this analysis.

Secondly, we also note that the crude output measures used for the VP sector may not capture case-mix differences very well. There may be a substantial difference in both time spent and number of therapists present for patients with chronic illnesses and patient with less serious and passing neurosis (Hatling and Magnussen, 1999). If the distribution of patients are uneven between clinics, this may effect our efficiency measures.

Thirdly, outpatient services are delivered by specialised personnel, e.g. physicians specialised as psychiatrists. psychologists specialised as clinical psychologists and nurses specialised as psychiatric nurses. In most cases, however, outpatient clinics are staffed with personnel undergoing training to become specialists. This implies that a substantial amount of time is spent on training, both by those undergoing it and by trained personnel acting as mentors. It is reasonable to assume that efficiency will be affected by the number of therapists engaged in some form of training. At present we have not included variables to adjust for this in our analysis, thus probably overestimating the potential for efficiency improvement.

- *There is constant returns to scale in the BUP sector, but variable returns to scale in the VP sector.*

We are not able to reject a hypotheses of constant returns to scale in the BUP sector. We note, however, that the clinics defining the best practice frontier are among the smallest clinics. Thus measurement for these clinics are likely to affect the results. For the VP sector the hypotheses of CRS is rejected. Initially we would expect that activity be proportional with staff. There might be however variations in other types of activity, in the sense that large clinics have a higher share of consultative work related to primary care and hospitals; and thus have a lower level of productivity.

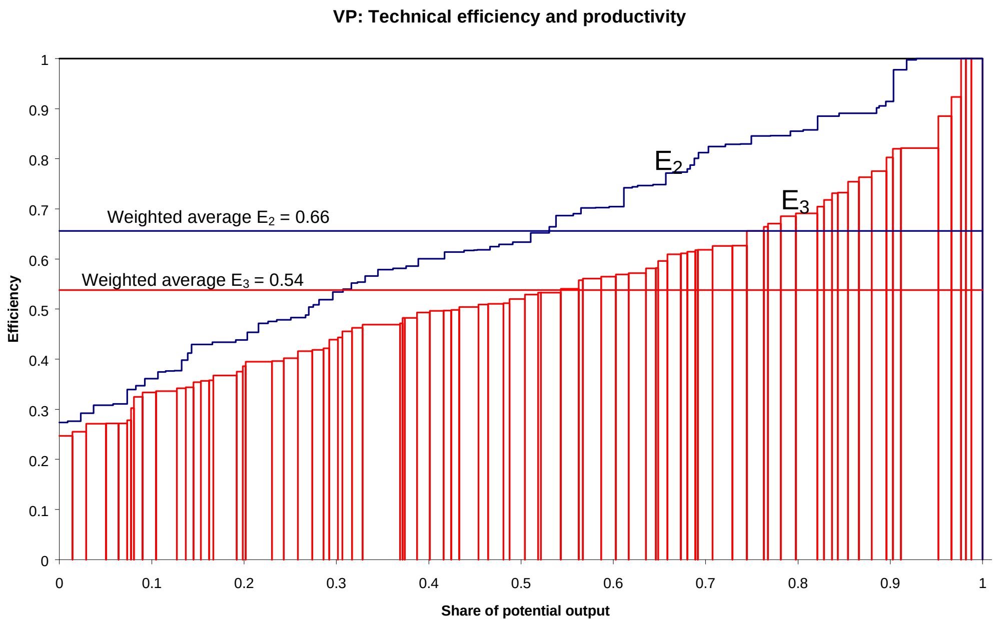


Figure 4: Distribution of technical efficiency and productivity estimates in the VP model.

- Staff composition does matter in both sectors, even though MPLs are quite similar.

To understand how staff composition could be expected to affect efficiency we need to look more closely at internal organisation of the outpatient clinics. For the moment sidestepping the fact that many will be in training position, there are broadly four types of therapeutic personnel in the clinics; psychiatrists, psychologists, nurses and social workers. In theory there is a division of labour between these professions. Social workers will, at the outset, have limited possibilities to perform individual therapy, psychiatrists are needed to administer medication, but will be less qualified to organise the patients living arrangements etc. In this respect the staffing mix would be a reflection of the patient mix of the clinic. What we observe in practice, however, is a production process where there is very little division of labour, and where specialised skills are utilised to transfer knowledge to other professions, rather than to use it in a clinical setting (Hatling & Magnussen, 1999).

Thus in many outpatient clinics the responsibility for treating a patient will be given to “first person available”, be it a social worker or a psychiatrist. In many cases the responsible person will perform tasks originally intended for another profession, in other cases he/or she will consult another member of the staff for advice.

In many ways this is a way of organising the activity that is inherently unproductive. Much time is spent on general staff meetings; both with respect to sorting out patients that are admitted and also with respect to discussing the treatment of individual patients. These meetings is a way of organising the treatment process that compensates for lack of knowledge on the therapist responsible for the patient, and work as a sort of internal education. On the other hand it is probably so that people in need of psychiatric care generally are better off when they can relate to fewer persons. Thus a model where the patient would meet 4-5 therapists during a treatment process could be even less productive than the model that is dominant today.

It is also worth noting that the unwillingness to utilise specialised skills by way of a more open division of labour is founded in a fundamental uncertainty about both how to diagnose and how to provide medical treatment for mental illnesses. In situations where there is uncertainty, each profession can “rightfully” maintain that it should be responsible for certain tasks. In the case of mental health services the professional disputes about who are/are not qualified to perform certain tasks have not been resolved; and the lack of specialisation is as much a result of this impasse as it is the result of a well conceived treatment concept.

7. Concluding comments

Measures of mental health illnesses are hard to find, and in this respect the analysis performed here should be treated with caution. Outcome measures such as the global assessment of functioning scores (GAF) will soon be available and will improve the policy value of this type of analysis.

References

Banker, R.D., 1993, Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273.

Banker, R.D., 1996, Hypothesis Tests Using Data Envelopment Analysis, Journal of Productivity Analysis 7, 139-159.

Banker, R.D., A. Charnes and W.W. Cooper, 1984, Some models for estimating technical and scale inefficiencies, Management Science 30, 1078-1092.

Bhattacharyya, G.K. and R.A.Johnson, 1977, Statistical concepts and methods (John Wiley & Sons, New York).

Charnes,A., W.W.Cooper and E.Rhodes, 1978, Measuring the efficiency of Decision Making Units, European Journal of Operational Research 2, 429-444.

Debreu, G., 1951, The coefficient of resource utilization, Econometrica 19, 273-292.

Farrell, M.J., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society 120, 449-460.

Färe, R., S.Grosskopf and C.A.K.Lovell, 1985, The measurement of efficiency of production (Kluwer-Nijhof, Boston).

Färe, R. and C.A.K.Lovell, 1978, Measuring the technical efficiency of production, Journal of Economic Theory 19, 150-162.

Førsund, F.R. and L. Hjalmarsson, 1974, On the measurement of productive efficiency, Swedish Journal of Economics 76, 141-54.

Førsund, F.R. and L. Hjalmarsson, 1987, Analyses of industrial structure: A putty-clay approach (Almqvist & Wiksell International, Stockholm).

Goodwin S (1997): Comparative Mental Health Policy. SAGE publications, London

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

Hagen (1998):

Halsteinli V (red) (1998): Psykiatritjenesten på 90 tallet. Rapport 5/98 SINTEF Norsk Institutt for Sykehusforskning, Trondheim

Hatling T og J Magnussen (1999): Arbeidsformer i psykiatriske poliklinikker. (kommer som Arbeidsrapport SINTEF Unimed NIS)

Johnson, N.L. and S.Kotz, 1970, Distributions in statistics: Continous univariate distributions-2 (Houghton Mifflin, Boston).

Kittelsen, S.A.C., 1999, "Monte Carlo simulations of DEA efficiency measures and hypothesis tests", Memorandum No. 9/1999, Department of Economics, University of Oslo.

Kneip, A., B.U.Park and L.Simar, 1996, A note on the convergence of nonparametric DEA efficiency measures, Discussion Paper 9603, Institut de Statistique, Université Catholique de Louvain.

Korostelev, A.P., L.Simar and A.B.Tsybakov, 1995a, Efficient estimation of monotone boundries, Annals of Statistics 23(2), 476-489.

Korostelev, A.P., L.Simar and A.B.Tsybakov, 1995b, On estimation of monotone and convex boundries, Publications de l'Institut de statistique de l'Université de Paris 39:, 3-18.

Magnussen, J., 1996, Efficiency measurement and the operationalization of hospital production, Health Services Research 31, 21-37.

SAMDATA, 1998,

Seiford, L.M., 1996, Data Envelopment Analysis: The evolution of the state of the art (1978-1995), Journal of Productivity Analysis 7, 99-137.

Simar, L. and P.W.Wilson, 1997, Nonparametric tests of returns to scale. Paper presented at the 5th European Workshop on Efficiency and Productivity Measurement, The Royal Veterinary and Agricultural University, Copenhagen.

Simar, L. and P.W.Wilson, 1998, Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, Management Science, 44, 49-61.

St:meld 25 (1996-97): Åpenhet og helhet. Om psykiske lidelser og tjenestetilbudene.

Valdmanis, V., 1992, Sensitivity analysis for DEA models. An empirical example using public vs. NFP hospitals. Journal of Public Economics, 48, 185-205.