THE UNIVERSITY of York

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

Data Envelopment Analysis Applied to Quality in Primary Health Care

Javier Salinas-Jiménez
Peter Smith

DISCUSSION PAPER 124

DATA ENVELOPMENT ANALYSIS APPLIED TO QUALITY IN PRIMARY HEALTH CARE

Javier Salinas-Jiménez[*] and Peter Smith[†]

* Departamento de Economía Aplicada, Universidad de Extremadura, Spain

† Department of Economics and Related Studies, University of York, United Kingdom

Address for correspondence:

Peter Smith
Department of Economics and Related Studies
University of York
York YO1 5DD
United Kingdom

Phone:	+44 904 433779 (from April 1995, +44 1904 433779)
Fax:	+44 904 433759 (from April 1995, +44 1904 433759)
E-mail:	pcs1@uk.ac.york.vax

July 1994

The Authors

Javier Salinas-Jiminez is Lecturer in the Department of Applied Economics, University of Extremadura, Spain. Peter Smith is Senior Lecturer in the Department of Economics and Related Studies, University of York.

Acknowledgements

We should like to thank Brenda Leese and Roy Carr-Hill of the Centre for Health Economics, University of York for their helpful comments.

Further Copies

Further copies of this document are available (at price £4.50 to cover the cost of publication, postage and packing) from:

The Publications Secretary
Centre for Health Economics
University of York
York YO1 5DD

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address, or telephone York (01904) 433648 or 433666.

ABSTRACT

The quality of primary care should ultimately be judged on the effect on health outcome of individual patients. However, for the foreseeable future, it is inconceivable that the necessary data will be available to implement this principle. And in any case, specification of the necessary statistical model is fraught with difficulty. This paper therefore applies data envelopment analysis (DEA) to quality in primary health care administration, in the belief that it offers a consistent and helpful "intermediate technology"for assessing performance. Many of the outputs of primary care are intangible and unfold over a long time period. It is therefore usual to judge the quality of primary care on the basis of process variables, which are particularly well suited to DEA. Moreover, DEA is not vulnerable to the misspecification bias that afflicts statistical models. The principal difficulty DEA gives rise to is the selection of relevant environmental variables. The issues are illustrated with an example from English Family Health Service Authorities.

INTRODUCTION

The UK National Health Service (NHS) seeks to provide comprehensive health care to all citizens, free at the point of delivery. It is a massive central government programme, accounting for £29.2 billion of state expenditure in England alone (5.5% of Gross Domestic Product). Of that total, £22.7 billion is spent on hospital and community services, £1.0 billion on central services, and £5.6 billion on family practitioner services, the subject of this paper (Department of Health, 1994). Family practitioners, or general practitioners (GPs), are responsible for delivering the bulk of primary care within the NHS.

Since 1991, the central government has devolved responsibility for local administration of primary care to 90 Family Health Service Authorities (FHSAs) (UK Government, 1990a). The members of the FHSAs are appointed by the Secretary of State for Health. The FHSAs are responsible for determining the geographical distribution of GPs, and for making contract payments to practices to cover staff salaries and running costs. FHSA allocations do not cover prescribing costs. Payments are based on population, with allowances for patient age and measures of deprivation. Since 1990 GPs have also received payments for achieving a variety of public health targets, relating to issues such as population surveillance, screening and immunization. GPs are technically independent contractors, and are not direct employees of the FHSAs. The FHSA therefore is in the somewhat awkward position of being responsible for the overall performance of primary care within its locality, but having only indirect control over the behaviour of GPs, the principal parties responsible for delivering services.

FHSAs are funded by the central government. The principal determinant of a FHSA's budget is its previous allocation, with incremental adjustments for perceived changes in circumstances. Before the creation of FHSAs there had been little attention paid to the quality of general practice, and the main concern of the central government had been with enforcing high standards of probity in the use of funds (Royal College of General Practitioners, 1985). However, a 1987 report signalled increased concern with efficiency, effectiveness and quality in primary care (Department of Health and Social Security, 1987), and there is now considerable interest in the extent to which FHSA budgets are being used efficiently to secure high quality primary care (Department of Health, 1993).

Quality in primary care is an elusive concept. Many of the important outcomes of GP intervention are subjective and materialize over very long time horizons (Harris, 1993). They are therefore resistant to satisfactory quantification. The recent trend in the UK, in line with the principles of the "Citizen's Charter", is to presume that the prime touchstone of quality should be client satisfaction (UK Government, 1990b). However, in the context of health services, there exist enormous problems in deriving useful measures of patient satisfaction (Carr-Hill, 1992). For example, expressions of satisfaction appear to be highly dependent on patient characteristics as well as GP behaviour (Hopton, Howie and Porter, 1993), and many of the important outcomes of primary care become evident only after many years of preventative intervention.

The usual approach to quality is therefore to presume that certain quantifiable features of the processes of primary care are highly correlated with "quality". Huntington (1993) characterizes low quality family practices as follows:

> "Such practices are typically very small, with lists of over 3,000 patients per doctor and minimal support staff. They achieve few or none of the 1990 contract targets, offer fragmented child health surveillance and contraceptive services, and no minor surgery. They are much akin to what in the United States are termed 'Medicaid mills'"

The UK Government now routinely publishes a large volume of data which seeks to address many of the issues raised by Huntington in the form of the annual "Health Service Indicators" package (National Health Service Management Executive, 1992). The purpose of this paper is to explore the extent to which data envelopment analysis (DEA) gives useful insights into FHSA performance as reflected in these data, and more generally to assess the usefulness of DEA in the primary health care sector. The next Section introduces the data to be used in the study. There follows a brief introduction to DEA, and to applications in the health sector. The results of this study are then reported, and the paper ends with some general conclusions.

THIS STUDY

In the absence of universally accepted direct measures of satisfaction, and given the impossibility of satisfactorily measuring outcome in primary care, it is inevitable that any examination of performance must resort to indirect proxies for quality, based on GP activity. There is however a surprising lack of data on GP activity patterns, and how they are related to patient satisfaction (Peter, Tate and Catchpole, 1989). There is clearly a need for research in this area (Harris, 1993). In this paper we focus attention on the performance of primary care within FHSAs by examining some readily available crude proxies for the quality of services in the knowledge that they may be highly imperfect and incomplete measures of quality. The judgements we come to are therefore highly conditional on how appropriate these proxies are. Nevertheless, we believe that the methodology we describe is generally applicable to the evaluation of quality in primary care, and we hope that this analysis might stimulate the search for better data.

The quality proxies are derived from the 1991/92 set of Health Service Indicators for FHSAs issued by the NHS Management Executive (1992). It proved possible to extract seven indicators of quality, along the lines suggested by Huntington, as follows:

- Y1 General medical practitioners per 10,000 patients on lists.
- Y2 The percentage of practices employing a practice nurse.
- Y3 The percentage of general medical practitioners who had a list of less than 2,500 patients.
- Y4 The percentage of general medical practitioners not practising single-handed.
- Y5 The percentage of general medical practitioners who had achieved the higher rate of payments for childhood immunization.
- Y6 The percentage of females aged 35 to 64, registered with the FHSA and who had an adequate cervical smear in the previous five and a half years.
- Y7 The percentage of practice premises which satisfied the minimum standards for facilities set out in the Statement of Fees and Allowances (paragraph 51.10, excluding practices exempt under paragraph 51.11).

Each of these variables is designed so that, other things being equal, a higher value suggests higher quality of care for patients. The first two variables are indicators of the availability of "front line" staff. Clearly, other things being equal, an increase in the number of general practitioners serving a fixed population (Y1) is likely to result in an increase in the perceived quality of local services. Residents are likely to have a greater number of practitioners from whom to choose, with a higher probability that they will find a practice offering the type of care they prefer.

The presence of a practice nurse (Y2) has been found to be a key indicator of the perceived quality of a practice. This is not simply because, other things being equal, it is likely to offer patients a more comprehensive primary care service. The presence of a practice nurse has also been found to be very highly correlated with other, unmeasured aspects of quality (Atkin *et al*, 1993).

Y3 is intended to capture the widely held belief that large list sizes may prevent GPs from giving patients the individual attention they require. Such "threshold" variables reflect the belief that, up to a certain point, the number of patients on a GPs list does not affect the quality of care they can offer. However, beyond a certain point (2,500 being the chosen level) it is considered that competition from other patients must begin to detract from the quality of care offered to a patient. Y3 is vulnerable to complications brought about by part time practitioners, and of course the 2,500 benchmark, although widely accepted, is arbitrary.

Y4 has been selected because it is believed that there are natural economies of scale in primary care. If this is the case, single handed practices may not be able to offer the range of services provided by larger practices. This issue has been the subject of intense scrutiny in the UK, where there is a preponderance of single-handed practitioners in inner cities (Royal College of General Practitioners, 1985). This is widely held to be a reliable indicator of poor quality of primary care in those areas, and to justify increased provision of hospital services in compensation.

Y5 and Y6 are process variables, reflecting the activities (but not necessarily the outcome) of primary care activity in an area. Clearly high immunization rates (Y5) are desirable in themselves, and may also indicate a more generally responsive and well organized system of primary care. The chosen measure of immunization levels is an indirect indicator, in the sense that it reflects the percentage of GPs achieving a certain level of coverage. Clearly a more direct and satisfactory measure would have been the percentage of children being immunized, but these data are not available. Takeup rates for cervical smears (Y6) may be more dependent on patient characteristics (Baker and Klein, 1991), but we seek to adjust for this in our choice of inputs (see below).

Finally, Y7 reflects poor quality of facilities, as defined by UK inspection standards. The factors taken into account are concerned predominantly with the physical condition of GP premises, including considerations such as wheelchair access, the privacy of consulting rooms, fire precautions and adequate waiting areas. Clearly this is another "threshold" variable, and suffers from the possibility of variations in interpretation around the country.

Most commentators would agree that the seven variables described above reflect important aspects of primary care. They might argue with the precise measures used, and might point to weaknesses in the quality of the data. They might also object that important aspects of quality have been omitted from the analysis - indeed that the measures described only give an indirect indication of quality, and do not address patient satisfaction, one of the prime benchmarks of quality. However, it is inevitable that, in the absence of markets or comprehensive surveys in which patients can express their preferences, bureaucracies must resort to measures of these sort in seeking to come to judgements about the quality of health care being delivered by their agencies. The purpose of this paper is to examine whether - given this inescapable position - DEA sheds any useful light on comparative performance.

There are of course numerous inputs which affect the ability of FHSAs to secure improvements in these outputs. Some of these inputs are under the direct control of the FHSA. Others reflect the characteristics of the locality, and are in general outside the control of the FHSA. Such "environmental" inputs might nevertheless be important determinants of FHSA performance. In this study, the following inputs are used:

X1 Gross expenditure on General Medical Services (in £'000s) per 10,000 FHSA resident population.

X2 Standardized Illness Ratio.

X3 Unemployment.

X1 reflects the resources under the control of the FHSA. They are used to administer the system, as well as to purchase general practitioner services. Raw expenditure has been adjusted using a social services cost adjustment factor to reflect higher factor costs in the south east of the country (UK Government, 1992). This adjustment is necessarily crude, but will always be required where there are significant cost differences between units being compared.

The remaining two inputs reflect uncontrollable circumstances. X2 is the ratio of observed to expected numbers reporting that they suffered from "limiting long-term illness" in the 1991 Census of Population. Expected numbers were calculated by applying age and sex specific English national average rates to the local population structure. X2 is therefore an index of local (self-reported) morbidity, a key determinant of demand for local GP services, and therefore of the quality of care that GPs may be able to deliver. X3 is the unemployment rate derived from the 1991 Census, and reflects more general deprivation in an area, which may also have important implications for the ability to deliver high quality primary care. Environmental factors of this sort will always have a profound impact on the efficiency of health programmes, and incorporation into DEA models is therefore essential.

Clearly it is possible to suggest countless additional environmental variables which affect the ability of FHSAs to secure high quality care. In particular, we have omitted consideration of the level of other health and welfare services in the FHSA. And it is possible to suggest numerous other indicators of deprivation. The wide extent of legitimate possible additions to the list of inputs highlights the importance of caution in interpreting the results of DEA. We would argue that any DEA application in such a complex field is likely to omit some important variables, and the extent to which this compromises the usefulness of the technique is discussed later in the paper.

Basic statistics for the data are shown in Table 1. Per capita expenditure varies widely, from £246 in the London FHSA of Greenwich, to £344 in Kensington and Chelsea, also in London. It is noteworthy that - because it is the lowest spending FHSA - Greenwich will be deemed efficient by DEA. Amongst the output variables, the Y5 (immunization) and Y6 (cervical smears) show the biggest variability in relation to median scores. The fact that some areas achieve a score of more than 100% for cervical cytology is an indication that the numerator (reported smears) is not a subset of the denominator (registered population), and that some women may have been tested more than once. This highlights a common weakness in epidemiological indicators.

	Median	Maximum	Minimum	Standard deviation
X1	298.38	343.62	245.63	22.16
X2	99.59	155.12	68.48	22.19
X3	8.93	21.12	5.10	3.41
Y1	5.09	7.89	4.50	0.43
Y2	88.43	100.00	54.55	10.11
Y3	89.29	100.00	69.35	7.42
Y4	87.24	100.00	65.12	7.68
Y5	66.32	96.93	9.93	21.01
Y6	71.11	112.37	4.33	17.68
Y7	95.09	100.00	30.65	12.01

Table 1: Summary statistics for inputs and outputs

Table 2 reports correlations between variables. The two environmental variables are highly correlated ($r = 0.80$), reflecting the acknowledged link between deprivation and morbidity. Most of the quality variables (outputs) are negatively correlated with the environmental inputs, confirming the generally accepted view that the quality of primary care is poorest in the more deprived areas. Spending (X1) is unrelated to deprivation, and only weakly correlated with many of the quality variables. The quality variables tend to be positively correlated with each other.

X1	1.00									
X2	-0.17	1.00								
X3	0.00	0.80	1.00							
Y1	0.48	-0.25	-0.26	1.00						
Y2	0.17	-0.20	0.67	-0.11	1.00					
Y3	0.42	-0.34	-0.44	0.67	0.52	1.00				
Y4	0.25	-0.28	-0.42	0.49	0.69	0.71	1.00			
Y5	0.28	-0.53	-0.73	0.45	0.65	0.65	0.68	1.00		
Y6	-0.13	0.00	-0.17	-0.03	0.26	0.06	0.13	0.22	1.00	
Y7	-0.17	0.02	-0.21	0.09	0.26	0.26	0.30	0.29	0.28	1.00
	X1	X2	X3	Y1	Y2	Y3	Y4	Y5	Y6	Y7

Table 2: Correlations between variables

Thus in this study we had available seven measures of quality of output and three inputs, of which two are environmental, or "uncontrollable" inputs. These data formed the framework for the DEA modelling that is described in the next section.

DATA ENVELOPMENT ANALYSIS

Data envelopment analysis is a linear programming technique for the estimation of the relative technical efficiency of a set of Decision Making Units (DMUs) producing a homogeneous set of outputs from common inputs (Charnes, Cooper and Rhodes, 1978). In the context of this study, the DMUs will be the 85 FHSAs for which a full dataset was available. The DEA model can be stated as follows. Let x_{ij} (i=1...m) be the m inputs used and y_{rj} (r=1...s) the s outputs produced by the DMU j (j=1...n). The technical efficiency of DMU 0 is then assessed as:

$$Max\ h_0 = \frac{\sum_{r=1}^{s} U_r y_{ro}}{\sum_{i=1}^{m} V_i x_{io}} \qquad (1)$$

subject to:

$$\frac{\sum_{r=1}^{s} U_r y_{rj}}{\sum_{i=1}^{m} V_i x_{ij}} \leq 1 \qquad j = 1 \ldots n.$$

$$U_r > 0, \quad V_i > 0 \quad \forall r, i.$$

That is, the DEA efficiency of DMU 0 is defined as the ratio of the weighted sum of the outputs to the weighted sum of inputs. The problem formulated above is to find the set of output weights U_r and input weights V_i for DMU 0 (the DMU that is being analyzed) which maximizes the apparent efficiency of DMU 0. The n restrictions require that, with the same set of weights, none of the other DMUs can have an efficiency score higher than one. If subject to this constraint it is possible to find a set of weights for which the efficiency ratio of DMU 0 is equal to one, DMU 0 will be regarded as efficient by DEA; otherwise it will be assigned an efficiency score less than one and will be regarded as inefficient.

Thus DEA offers a measure of comparative efficiency by comparing an FHSA with its peers for a given set of inputs and outputs. The weights U_r and V_i can be interpreted as the relative prices of outputs and inputs, and so the ratio being maximized is a ratio of the benefits produced by a DMU to the costs it incurs. The unique feature of DEA is that the "prices" are not chosen by the analyst, but are selected by the linear programme to show the FHSA being examined in the most favourable possible light. Thus - if inputs and outpus have been correctly specified - the estimate of efficiency generated by DEA is often conservative, in the sense that an unrealistic set of prices might have been chosen. Moreover, some FHSAs may be deemed efficient simply because there were no peers with which to compare them.

One of the by-products of DEA is that - for units it deems inefficient - it produces a set of efficient peers with which the apparently inefficient unit is being compared. The comparison is formed by taking a weighted average of each of inputs and outputs of the efficient units. The performance of the "composite" FHSA formed by this weighting procedure gives achievable targets for the inefficient unit.

The DEA model has some important advantages over parametric and econometric approaches. Two of the most important are, firstly, that it does not impose a particular functional form on the production frontier; and secondly, it readily enables the user to handle multiple-output, multiple-input technologies, a feature which is especially important in the assessment of efficiency in public sector activities.

It is nevertheless worth noting some limitations of the technique. Firstly, DEA is a deterministic technique. It relies on identifying outliers (the most efficient units), in contrast to techniques such as regression analysis, which average out stochastic errors. Thus, any data errors might lead to seriously misleading conclusions. Secondly, the results provided by DEA might be very sensitive to the number of variables (inputs and outputs) included in the analysis. As explained by Thrall (1989) the number of DMUs assessed as efficient units and the efficiency scores set for each unit may increase and cannot decrease when new variables are included in the analysis. Therefore, if the number of DMUs is small compared to the number of factors considered in the efficiency assessment, the DEA approach may lead to substantial overestimates of the efficiency of DMUs. It has been suggested, as a general rule of thumb, that it is advisable to ensure that the number of DMUs is at least three times the combined number of inputs and outputs (Banker, Charnes, Cooper, Swarts and Thomas, 1989).

Thirdly, as noted above, extreme units may be regarded as efficient units only because they are not comparable with the rest of the units included in the sample. This is a consequence of the way in which prices are selected in DEA. Each DMU chooses the criteria by which it wishes to be judged, selecting the prices that show the unit in "its most favourable light". If this is considered unacceptable, price variations may be restricted (Wong and Beasley,1990; Thompson, Langemeier, Lee, Lee and Thrall, 1990; Kornbluth, 1991; Roll and Golany, 1993).

Finally, the DEA approach only yields a measure of *relative* technical efficiency. Providing that the model is correctly specified (has a complete set of inputs and outputs), a DMU assessed as inefficient by DEA is intrinsically inefficient. A unit assessed as efficient by DEA is technically efficient, *given the practices observed in the sample being analyzed.* Obviously, the possibility of greater technical efficiency than the observed in the sample cannot be ruled out. As pointed out by Ganley and Cubbin (1992), DEA-efficiency does not necessary imply Pareto-efficiency.

In spite of the burgeoning literature on DEA (Ali and Seiford, 1993), there has been a comparatively modest published research effort in the health sector, and most health studies have been concerned with the efficiency of hospital services. Early examples include an exploratory study by Sherman (1984), who used DEA to examine the efficiency of seven teaching hospitals in Massachusetts. Sexton *et al* (1989) and Burgess and Wilson (1993) have examined Veterans Administration hospitals, while Grosskopf and Valdmanis (1987), Ozcan, Luke and Haksever (1992) and Valdmanis (1992) have used DEA to explore the impact of ownership type on efficiency. Other studies in the hospital sector include Banker, Conrad and Strauss (1986), Register and Bruning (1987) and Ozcan and Luke (1993).

Other health-related DEA studies include routine nursing services (Nunamaker, 1983); rural primary health care programmes (Huang and McLaughlin, 1989); public health services (Pina and Torres, 1992); and maternity services (Thanassoulis, Boussofiane and Dyson, 1991). Färe, Grosskopf, Lindgren and Roos (1992) examine panel data relating to the efficiency of pharmacies.

RESULTS

It must be emphasized that the example used here is illustrative, and is not intended to pass any sort of definitive judgement on English FHSAs. It seeks to identify areas in which - using published data sources - there is *prima facie* evidence of scope for improved performance. The DEA terms "efficient" and "inefficient" should therefore be viewed with some caution, as they refer to *relative* levels of efficiency, given a particular set of assumptions.

Hitherto, in analysing data of the sort described above, the preoccupation of the UK Department of Health has been with constructing "league tables" of performance on each individual indicator. Indeed, the data are distributed in computer files which contain each authority's rank on an indicator, as well as the raw data. There has been an attempt to encourage users to explore the data using a rudimentary expert system (Bowen and Payling, 1987), but this has not been very widely used. Thus, if they are used at all, the data are usually subjected to a crude univariate analysis, and the trade-offs between variables are not modelled.

As a prelude to the DEA work, we sought using statistical means to estimate a traditional cost function, in line with standard economic theory. Table 3 reports the ordinary least squares regression of expenditure (X1) on the remaining input and output variables which showed the highest adjusted R^2 (= 0.364). Both environmental variables (X1 and X2) show statistically significant effects, but the coefficient on illness (X2) has an unexpected sign. With the exception of inspection standards (Y7), all the signs on outputs are positive. The results in Table 3 are however probably misleading, because the relationship between expenditure, environment and quality is likely to be highly complex, involving some element of simultaneity between the variables. This being the case, the ordinary least squares regression is almost certainly misspecified, and in order to build a satisfactory statistical model, it is probably necessary to specify a system of equations, and to estimate them using more advanced statistical techniques. In the absence of any good understanding of the processes whereby epidemiology, health services and health status are related, this is a daunting task.

Variable	Coefficient	Standard error
X2 Illness	-0.453	0.151
X3 Unemployment	4.965	1.163
Y3 List size < 2,500	1.286	0.322
Y5 Immunization	0.372	0.134
Y7 Premises	-0.482	0.176
Constant	201.68	31.03

Table 3: Regression of per capita expenditure (X1) on inputs and outputs

The case for using DEA in these circumstances is that it is not necessary to specify a comprehensive model in order to obtain meaningful results. Being a deterministic frontier method, DEA requires no specification of functional form, and - more importantly in this case

- sidesteps the problem of simultaneity that bedevils statistical methods.

In implementing DEA, note that all the variables here are ratios which have been adjusted for the size of the FHSA. Underlying the analysis, therefore, is an assumption of constant returns to scale in FHSA activities. DEA uses weighted averages of observations on individual units to form the efficient frontier. This gives rise to two difficulties. First, the datum obtained from a weighted average of ratios is not in general the same as the datum obtained by weighting numerator and denominator separately (Fernandez-Castro and Smith, 1994). This problem is unlikely to be important in this study. And second, comparison should be by interpolation between observations only. Extrapolation between observations cannot be permitted. Rather confusingly, this criterion is satisfied by invoking what Banker, Charnes and Cooper (1984) refer to as the variable returns to scale constraint. It is used in this study simply to inhibit exptrapolation.

Note also that the implementation of DEA reported here requires that outputs are held constant, and therefore indicates the extent to which inputs can be reduced without reducing any of the outputs. The fact that X2 (illness) and X3 (unemployment) are uncontrollable inputs therefore requires that they are treated as such in the analysis, along the lines suggested by Banker and Morey (1986). The analysis then indicates the extent to which X1 (expenditure) can be reduced subject to the constraint that comparators suffer equal or worse environmental circumstances (X2 and X3).

The DEA model was run first using all ten factors described above. Missing values required that only 85 of the 90 DMUs were used in the analysis. This model resulted in 43 (51%) of the FHSAs being deemed efficient and 42 (49%) inefficient. Amongst the inefficient units, the average level of DEA efficiency was 0.926. Thus, using the full data set, there was only limited evidence of any significant potential for efficiency savings. However, it should be remembered that the results have been secured allowing total weight flexibility. Thus an FHSA might secure high efficiency by emphasizing only one or a few dimensions of performance. Further work might examine the impact of imposing restrictions on the flexibility of weights.

Kensington & Chelsea	0.7323
Brent & Harrow	0.7384
Enfield & Haringey	0.8329
Ealing Hammersmith Hounslow	0.8359
Wolverhampton	0.8488
Sandwell	0.8555
Solihull	0.8786
East Sussex	0.8951
Walsall	0.8961
Humberside	0.8972
Leicestershire	0.8979
Kirklees	0.9058
Lancashire	0.9092
Stockport	0.9106
Salford	0.9107
Lincolnshire	0.9207
Avon	0.9216
Leeds	0.9288
Wirral	0.9311
Warwickshire	0.9320
Merton Sutton Wandsworth	0.9369
Sefton	0.9410
Barking & Dagenham	0.9480
Gateshead	0.9493
Bedfordshire	0.9541
Bolton	0.9545
Cambridgeshire	0.9578
West Sussex	0.9592
Kingston & Richmond	0.9608
Essex	0.9623
Shropshire	0.9651
Bury	0.9659
Oxfordshire	0.9734
Trafford	0.9750
Berkshire	0.9775
Redbridge	0.9776
Staffordshire	0.9851
Dorset	0.9873
Tameside	0.9882
Nottingham	0.9891
Hampshire	0.9962
Hillingdon	0.9973

Table 4: FHSAs deemed inefficient by DEA

The FHSAs deemed inefficient are shown, together with their relative efficiency scores, in Table 4. Using this analysis, the least efficient authority is Kensington and Chelsea, which had an efficiency of 0.732. The use of DEA therefore implies that this FHSA could reduce its expenditure by 26.8% if it performed as well as its reference group, which comprised a weighted average of Sunderland (28%), Barnsley (9%) and Greenwich (63%).

The remaining FHSAs were deemed 100% efficient. However, as noted above, this may be due to lack of comparability rather than true efficiency. Thus, for example, Barnsley has the highest level of limiting illness in the sample (X2 = 155.12). This FHSA will therefore always be deemed efficient if this factor is included in the analysis because it suffers uniquely adverse environmental circumstances, and so cannot be compared with any other FHSA (Norman and Stoker, 1991). As an indication of *robustly* efficient FHSAs, therefore, Table 5 shows those efficient FHSAs which formed part of the comparison group of more than 10 inefficient FHSAs.

North Tyneside	26
Croydon	23
Isle of Wight	16
Greenwich	16
Sunderland	14
Bromley	13
Barnsley	10
Wiltshire	10
Durham	10

Table 5: Efficient FHSAs occuring in reference set of at least 10 inefficient FHSAs

The importance of incorporating environmental factors into the analysis can be illustrated by rerunning the analysis without illness (X2) or unemployment (X3). The effect of omitting a salient input in DEA is to constrain the associated weight in (1) to be zero. This implies that no efficiency score can be increased by the omission, and some might be reduced. In fact, the biggest reduction in measured efficiency brought about by running this reduced model is experienced by the City and East London FHSA, which was fully efficient under the original model, and which is now deemed to have an efficiency of 0.770. Scrutiny of its environmental variables indicates high levels of illness (134.76) and unemployment (20.9), illustrating the great sensitivity of results to choice of environmental variables amongst disadvantaged areas.

DISCUSSION

In principle, there is only one completely intellectually coherent way to assess the impact of administrative units on outcome in health care. This entails deriving suitable measures of outcome amongst *individuals*, and applying multilevel modelling techniques to infer the impact of the institution on individual outcome. This approach has enjoyed considerable success in the education sector, where clients are readily identifiable, where their calibre on entry is known with some precision, and there is some measure of outcome (examination success) (Paterson and Goldstein, 1991). However, in spite of some preliminary efforts (Gatsonis, Normand, Liu and Morris, 1993), research in the health sector is at a much less advanced stage, and the much greater complexity of the underlying model suggests that transferring multilevel techniques from the education to the health sector will be a considerable challenge. In any case, there is a notorious shortage of individual health data, and for the foreseeable future the performance analyst will have available only aggregate data with which to assess the performance of health administrations.

This being the case, there is a limited armoury of available techniques. Clearly piecemeal univariate examination of single indicators is likely to be inadequate, given the multidimensional nature of health outcome. Multivariate statistical methods have therefore enjoyed considerable popularity. However, they suffer from a number of weaknesses:

- the need to specify a parametric functional form;
- the difficulty of handling multiple inputs and inputs;
- the need to examine the error structure for possible simultaneity;
- the need for large sample sizes.

Most fundamentally, conventional statistical methods are compromised by the poorly understood nature of the processes whereby epidemiology and health care interact to yield health outcomes. Data envelopment analysis therefore offers an attractive possibility as a device intermediate between crude univariate analysis and the unmanageably complex statistical techniques. The purpose of this paper has been to assess the strengths and limitations of DEA applied to the health sector, using the primary care case study as an example.

In a survey of OR applied to health services, Rosenhead (1978) pointed to the dangers of applying a traditional OR approach to health services planning. His conclusion was that "we should prize approaches which: make reduced demands on data; reject optimisation in favour of coordination; accept uncertainty and try to keep options open; are not restricted to hierarchical deduction, but facilitate participation; [and] do not attempt a technocratic abolition of politics".

In many respects, DEA is sympathetic to the criteria set out by Rosenhead. It offers some insight into performance, even with a limited dataset, and a small number of observations. In allowing weight flexibility, it does not insist that there is a single way of being efficient, and respects individual DMU choices regarding the importance of outputs. In the same spirit, the flexibility with which inputs and outputs can be selected, and the flexibility attached to valuations of outputs, respects the heterogeneity of political views relating to the health sector.

The weakest feature of DEA as applied to health is the treatment of uncertainty. A previous study by the authors of this paper sought to apply DEA to maternity services. It was abandoned because the outcome measures (various mortality and morbidity data) were highly susceptible to random fluctuation, leading to a frontier determined by "lucky" DMUs which might have been - on average - impossible to achieve. This phenomenon will always arise when DMUs are assessed on outcome measures determined by a relatively small number of very important events (such as deaths). In this respect, theoretical developments in stochastic DEA might be particularly important for the health sector (Sengupta, 1987).

This study has been less ambitious, in that the outcome measures can only be inferred by proxies based on routine process. In these circumstances, the DEA results are likely to be relatively stable, and the targets achievable. The principal limitations identified in this study are the crucial importance of environmental variables, and the difficulty of deciding which environmental inputs to include in the absence of any statistical tests of their importance in influencing outcome. As shown in Smith (1993), omission of a relevant input can lead to serious biases in efficiency estimates. As a result, it is unlikely that DEA can be used in isolation from results from conventional statistical analysis, which are needed to inform the choice of inputs. Nevertheless, the study presented here does suggest that DEA offers a useful "intermediate technology" for assessing performance in health care, when output measures relate to relatively routine functions, and when there is some *a priori* basis on which to select environmental inputs.

REFERENCES

Ali, A. I. and Seiford, L.M. (1993), "The mathematical programming approach to efficiency analysis", in Fried, H. O., Lovell, C. A. K. and Schmidt, S. S. (1993), *The measurement of productive efficiency*, Oxford: Oxford University Press.

Atkin, K., Lunt, N., Parker, G. and Hirst, M. (1993), *Nurses count: a national census of practice nurses*, York: Social Policy Research Unit.

Baker, D. and Klein, R. (1991) "Explaining outputs of primary health care: population and practice", *British Medical Journal*, 303(6796), 225-229.

Banker, R.D. Charnes, A. and Cooper, W.W. (1984), "Some models for estimating technical and scale inefficiencies in data envelopment analysis", *Management Science*. 30(9).

Banker, R.D., Charnes, A., Cooper, W.W., Swarts, J. and Thomas, D. (1989), "An introduction to data envelopment analysis with some of its models and their uses", in *Research in Governmental and Nonprofit Accounting*, eds J.L. Chan and J.M. Patton, Greenwich: JAI Press.

Banker, R.D., Conrad, R. and Strauss, R. (1986), "A comparative application of data envelopment analysis and translog method: an illustrative study of hospital production", *Management Science*. 32(1), 30-44.

Banker, R.D. and Morey, R.C. (1986), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research*, 34(4), 513-521.

Bowen, B. and Payling, L. (1987), "Expert systems for performance review", *Journal of the Operational Research Society*, 38, 929-934.

Burgess, J. F. and Wilson, P. W. (1993), "Technical efficiency in Veterans Administration Hospitals", in Fried, H. O., Lovell, C. A. K. and Schmidt, S. S. (1993), *The measurement of productive efficiency*, Oxford: Oxford University Press.

Carr-Hill, R. A. (1992) "The measurement of patient satisfaction", *Journal of Public Health Medicine*, 14, 236-249.

Charnes, A., Cooper, W. and Rhodes, E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research*, 2(6), 429-444.

Department of Health (1993) *GP Practice Charters: making them happen*, London: Department of Health.

Department of Health (1994) *Departmental Report*, London: HMSO.

Department of Health and Social Security (1987), *Promoting Better Health - the government's programme for improving primary health care*, (Cm 249), London: The Department.

Färe, R., Grosskopf, S., Lindgren B. and Roos P. (1992), "Productivity changes in Swedish pharmacies 1980-1989: a non-parametric Malmquist approach", *Journal of Productivity Analysis*, 3, 85-101.

Fernandez-Castro, A. and Smith, P. (1994), "Towards a general non-parametric model of corporate performance", *Omega: International Journal of Management Science*, 22(3).

Ganley, J.A. and Cubbin, J.S. (1992), *Public sector efficiency measurement: applications of data envelopment analysis*, Amsterdam: North Holland.

Gatsonis, C., Normand, S. L., Liu, C. and Morris, C. (1993), "Geographic variation of procedure utilization: a hierarchical model approach", *Medical Care*, 31, 54-59.

Grosskopf, S. and Valdmanis, V. (1987), "Measuring hospital performance - a non-parametric approach", *Journal of Health Economics*, 6, 89-107.

Harris, A. (1993), "Developing a research and development strategy for primary care", *British Medical Journal*, 306, 189-192.

Hopton, J.L., Howie, J.G. and Porter, A.M. (1993) "The need for another look at the patient in general practice satisfaction surveys", *Family Practice*, 10, 82-87.

Huang, Y.L. and McLaughlin, C.P. (1989), "Relative efficiency in rural primary health care: an application of data envelopment analysis", *Health Services Research* 24 (2), 143-158.

Huntington, J. (1993), "From FPC to FHSA to ... health commission?", *British Medical Journal*, 306, 33-36.

Kornbluth, J.S.H. (1991), "Analyzing policy effectiveness using cone restricted data envelopment analysis", *Journal of Operational Research Society*, 42(12), 1097-1104.

National Health Service Management Executive (1992), *Health Service Indicators Handbook*, London: The Executive.

Norman, M. and Stoker, B. (1991), *Data envelopment analysis: the assessment of performance*, Chichester: Wiley.

Nunamaker, T.R., (1983), "Measuring routine nursing service efficiency: a comparison of cost per patient day and data envelopment analysis models", *Health Services Research*, 18, 183-205.

Ozcan, Y. A. and Luke, R. D. (1993), "A national study of the efficiency of hospitals in urban markets", *Health Services Research,* 27, 719-739.

Ozcan, Y. A., Luke, R. D. and Haksever, C. (1992), "Ownership and organizational performance: a comparison of technical efficiency across hospital types", *Medical Care*, 30, 781-794.

Paterson L and Goldstein H (1991) New statistical methods for analysing social structures: an introduction to multilevel models, *British Educational Research Journal*, 17(4), 387-393.

Peter, L., Tate, J.B. and Catchpole, P. J. (1989) "Practice activity analysis: collaboration between general practitioners and a Family Practitioner Committee", *Journal of the Royal College of General Practitioners*, 39, 297-299.

Pina, V. and Torres, L. (1992), "Evaluating the efficiency of nonprofit organizations: an application of data envelopment analysis to the public health services", *Financial Accountability and Management*, 8(3), 213-225.

Register, C.A. and Bruning, E.R. (1987), "Profit incentives and technical efficiency in the production of hospital care", *Southern Economic Journal*, 53, 899-914.

Roll, Y, and Golany, B, (1993), "Alternate methods of treating factor weights in DEA", *International Journal of Management Science*, 21(1), 99-109.

Rosenhead, J. (1978), "Operational research in health services planning", *European Journal of Operational Research*, 2, 75-85.

Royal College of General Practitioners (1985), *Quality in general practice: policy statement 2,* London: The College.

Sengupta, J. K. (1987), "Data envelopment analysis for efficiency measurement in the stochastic case", *Computers and Operations Research*, 14(2), 117-129.

Sexton, T.R., Leiken, A.M., Nolan, A.H., Liss, S. Hogan, A. and Silkman, R.H. (1989), "Evaluating managerial efficiency of Veterans Administration Medical Centers using data envelopment analysis", *Medical Care.* 27(12), 1175-1188.

Sherman, H.D. (1984), "Hospital efficiency measurement and evaluation - empirical test of a new technique", *Medical Care.* 22(10), 922-938.

Smith, P. (1993), "Misspecification bias in data envelopment analysis", *Discussion Paper 93/15*, Department of Economics, University of York.

Thanassoulis, E., Boussofiane, A. and Dyson, R. G. (1991), "DEA models for the assessment of the provision of perinatal care in England on the basis of quantity and quality of outputs", *Warwick Business School Research Papers 26.* Coventry: University of Warwick.

Thompson, R.G., Langemeier, L.N., Lee, C., Lee, E, and Thrall, R.M, (1990), "The role of multiplier bounds in efficiency analysis with application to Kansas farming", *Journal of Econometrics*, 46, 93-108.

Thrall, R.M. (1989), "Classification transitions under expansion of inputs and outputs in DEA", *Managerial and Decision Economics*, 10(2), 159-162.

UK Government (1990a), *Working for Patients*, (Cm 555), London: HMSO.

UK Government (1990b), *The Citizen's Charter*, (Cm 1599), London: HMSO.

UK Government (1992), *The Revenue Support Grant Distribution (Amendment) (No 2) Report (England)*, London: HMSO.

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

Wong, Y.H.B, and Beasley, J.E, (1990), "Restricting Weight Flexibility in Data Envelopment Analysis", *Journal of Operational Research Society*, 41(9), 829-835.