

ESTUDOS I



FACULDADE de ECONOMIA da UNIVERSIDADE do ALGARVE

ESTUDOS I

Cidadania, Instituição e Património

Economia e Desenvolvimento Regional

Finanças e Contabilidade

Gestão e Apoio à Decisão

Modelos Aplicados à Economia e à Gestão

A Faculdade de Economia da Universidade do Algarve



Faculdade de Economia da Universidade do Algarve

2004

COMISSÃO EDITORIAL

António Covas
Carlos Cândido
Duarte Trigueiros
Efigénio da Luz Rebelo
João Albino da Silva
João Guerreiro
Paulo M.M. Rodrigues
Rui Nunes

FICHA TÉCNICA

Faculdade de Economia da Universidade do Algarve

Campus de Gambelas, 8005-139 Faro
Tel. 289817571 Fax. 289815937
E-mail: ccfeua@ualg.pt
Website: www.ualg.pt/feua

Título

Estudos I - Faculdade de Economia da Universidade do Algarve

Autor

Vários

Editor

Faculdade de Economia da Universidade do Algarve
Morada: Campus de Gambelas
Localidade: FARO
Código Postal: 8005-139

Compilação e Design Gráfico

Susy A. Rodrigues

Revisão de Formatação e Paginação

Lídia Rodrigues

Fotolitos e Impressão

Serviços Gráficos da Universidade do Algarve

ISBN

972-99397-0-5 - Data: 26.10.2004

Depósito Legal

218279/04

Tiragem

500 exemplares

Data

Novembro 2004

**RESERVADOS TODOS OS DIREITOS
REPRODUÇÃO PROIBIDA**

Exploring the use of DEA for formative evaluation in primary care: An application to England

Carla A. F. Amado

Faculdade de Economia, Universidade do Algarve

Resumo

Este artigo discute a avaliação do desempenho formativa no contexto da prestação de cuidados primários.

Na primeira parte do artigo, discutimos o contexto do trabalho: a prestação de cuidados primários em Inglaterra. Nesta introdução discutimos também os objectivos deste estudo.

Na segunda parte do artigo fazemos uma discussão da metodologia proposta para medir a eficiência e a efectividade organizacional. A técnica não paramétrica de análise de dados conhecida por Data Envelopment Analysis (DEA) é introduzida.

Na terceira parte do artigo fazemos uma revisão da literatura nesta área. Os estudos publicados que usam o DEA para comparar centros de saúde são revistos de uma forma crítica. Esta revisão da literatura permite identificar as deficiências dos estudos até agora levados a cabo nesta área, no sentido de elaborar um novo framework para comparação de centros de saúde.

Este framework é apresentado na quarta parte do artigo e tem como objectivo estabelecer uma ligação entre as necessidades locais de saúde, os recursos usados, os serviços prestados e os resultados atingidos. Com base nas relações estabelecidas entre estes elementos, definimos quatro critérios para a avaliação do desempenho na prestação de cuidados de saúde primários: equidade na utilização dos serviços; eficiência na utilização dos recursos, efectividade clínica e efectividade de custos na prestação de serviços. Embora o framework desenvolvido seja específico para a área da saúde, pequenas alterações ao nível da terminologia são suficientes para o adaptar para a avaliação de desempenho na prestação de outros serviços públicos.

Palavras-chave: Data Envelopment Analysis, Avaliação de desempenho formativa, Cuidados de saúde primários.

Abstract

This paper discusses the formative evaluation of primary health care providers.

In the first part of the paper, we discuss the context of study: primary health care delivery in England. In this introduction we also discuss the objectives of this study.

In the second section, we discuss the methodology used: Data Envelopment Analysis (DEA). This non-parametric technique was proposed to measure efficiency, service effectiveness and cost effectiveness, given the fact that it can handle multiple inputs and outputs without requiring the specification of a functional form for the production function.

In the third section, we present a critical literature review of the studies that have compared primary care providers using DEA. This literature review allows us to

identify the shortcomings of these studies, in order to identify the aspects that need to be addressed in the development of a new framework.

In the fourth section, we discuss the development of a new conceptual framework to compare primary care providers. This framework was developed in collaboration with two primary care groups and two primary care trusts in England and aims to establish a

link between the local needs, the resources used, the services delivered and the outcomes achieved in primary care. Based on the relationships between these elements, four performance assessment criteria were defined: equity, efficiency, service effectiveness and cost effectiveness. Despite the fact that this framework has been developed for the specific area of health care, little changes in the terminology would be sufficient in order to apply this framework to the formative evaluation in the delivery of other public services.

Keywords: Data Envelopment Analysis, Formative evaluation, Primary health care delivery.

1. Introduction

1.1. Introduction

Scarcity of public resources and increasing health care needs point towards the importance of constantly looking for better ways of delivering and managing health care services. Performance assessment in health care delivery is seen as a method of ensuring that taxpayers' money is spent in the best possible way. As a result, performance assessment of health care systems is currently the subject of major interest, both nationally and internationally. Since 1983, when the first set of Performance Indicators (PIs) were introduced in the National Health Service (NHS) in the UK, calls for new ways to measure performance in health care delivery have not abated. From hundreds of PIs, numerous league tables have been produced and, despite continuous protest from health care professionals, the media uses the numbers to produce sensational headlines.

In the NHS the importance given to performance assessment is clear, as means to achieve care of uniformly high standards (The Department of Health 1997). Having recognised the limitations of the previous system of operation, management and control in place during the Conservative Government (in particular, a disregard for primary care management and equity considerations), the Labour Government has developed what they call 'a third way' of managing and delivering public services: a system based on partnership and driven by performance (The Department of Health 1997). A new framework for performance assessment (PAF) has been developed. In this framework six performance assessment criteria have been identified: (1) health improvement; (2) fair access to health services; (3) effective delivery of appropriate health care; (4) efficiency; (5) patient/ carer experience and (6) health outcomes of the NHS care. Structures and agencies have also been introduced to guarantee that quality is reviewed and improved. The National Institute for Clinical Excellence and National Service Frameworks (NSFs) set the standards based on clinical evidence and cost effectiveness. The Commission for Health Improvement monitors the achievement of these standards. Furthermore, at a local level, each organisation and professional is

also responsible for continuous improvement of quality in clinical practice. The objective is to make ‘best practice the norm’.

For the first time, primary care has been characterised as central to the development of the ‘*new NHS*’ (The Department of Health 1997). Having recognised its importance, the Government implemented structural reforms regarding primary care provision, management and accountability.

The reforms initiated in 1997 appear to be based on three basic assumptions regarding the delivery and management of public services. Firstly, that the methods and models successful in the private sector can also lead to success in the public sector. In this respect, the emphasis is “on the development of more flexible, responsive and entrepreneurial forms of conduct which would overcome the assumed stasis, rigidity and inefficiency of ‘bureaucracy’” (Du Gay, 2000: 63). Secondly, that performance in primary health care can be appropriately measured. All that is required is a wide list of performance indicators (PIs) to collect data to support each one of them and then to develop a weighting system in order to rank organisations and identify *best practice*. Thirdly, that efficiency, effectiveness, fairness of access and patients’ responsiveness can be improved simultaneously. Altogether, it is assumed that as long as several dimensions of performance are outlined, regularly monitored and as long as clear incentives are in place, it is possible to identify and reward best practice, leading to its reproduction across the country.

However, only a few studies have attempted to compare the performance of several primary health care providers in order to provide some guidance regarding the validity of these assumptions in this context. In this paper we aim to contribute to a discussion regarding appropriate ways to compare the performance of primary care providers and the effective ways to use these results in order to improve performance in primary care.

We developed a conceptual framework for performance assessment in primary care, in collaboration with newly established Primary Care Groups and Trusts (PCG/Ts). It was our objective to use this framework to compare a sample of General Practitioners (GP) surgeries in terms of their delivery of primary care as a whole. However, considering the Government’s strategy of moving towards speciality based delivery and performance assessment, and data availability, it became necessary to focus on a speciality: diabetes care. We propose the following criteria for evaluation: equity, efficiency, service effectiveness and cost effectiveness. These criteria and the whole framework for evaluation were developed in collaboration with the participating PCG/Ts and validated with the participating surgeries.

1.2. Research objective

This research study had the following objective: to develop a comprehensive performance assessment framework for formative evaluation in primary care, that encompasses the following aspects:

- The existence of multiple, incommensurable and often conflicting values underlying public service delivery. This points towards the development of a framework that takes into account the trade-offs between the different

criteria. We are particularly interested in studying the relationships between equity, efficiency and effectiveness;

- The intrinsically different *nature* of primary care provision when compared with other forms of service delivery. Primary care provision involves a relationship established between the patient and the professional, which relies on *trust* and is characterised by a strong asymmetry of knowledge and information. This means that performance assessment of primary care providers is necessarily partial, because there are essential elements that are very difficult to measure, such as the quality of the diagnosis;
- The difficulties in measuring the ultimate outcomes of primary care delivery;
- The existence of many non-controllable factors impacting on the outcomes of primary care, such as area deprivation and patient compliance;
- The potential for responsive dysfunctional behaviour to any performance assessment exercise. This is particularly relevant in primary health care because of the asymmetry of information and the partiality of the assessment.

1.3. Outline of this paper

This introductory section provided an overview of this research study and of its objectives.

Section 2 provides an introduction to Data Envelopment Analysis (DEA). Firstly, we discuss the basic DEA model developed by Charnes, Cooper and Rhodes (1978). This is followed by a discussion of cost efficiency and allocative efficiency, and a method to calculate it by using a DEA model.

Section 3 reviews the relevant literature. We review the studies that have been undertaken using DEA to compare primary care providers. By reviewing the performance assessment studies in primary health care, we identify the main issues that still need to be addressed.

Section 4 discusses the development of a conceptual framework for formative evaluation of GP surgeries. This conceptual framework aims to establish a link between the local needs, the resources used, the services delivered and the outcomes achieved. The difficulties encountered in applying this conceptual framework to evaluate primary care as a whole are discussed and the decision to focus on a speciality is justified. The advantages and limitations of comparing primary care providers based on a speciality are also discussed.

Section 5 concludes the paper and outlines the main contributions of the study. The limitations of this study and suggestions for further research in this area are also included.

2. Methodology: An introduction to data envelopment analysis

2.1 Introduction

In this section we introduce the technique proposed to compare primary care providers: DEA. We begin with a discussion of the measurement of technical efficiency by using the standard constant returns to scale model known as the CCR model (Charnes, Cooper and Rhodes 1978). This is followed by a discussion of the measurement of cost efficiency. The decomposition of cost efficiency into technical and allocative components is also presented.

2.2 Using DEA to measure efficiency

Since it was first proposed by Charnes, Cooper and Rhodes (1978), DEA has been widely used to compare the efficiency of homogeneous Decision Making Units (DMUs) that use a set of inputs to produce a set of outputs. This technique calculates efficiency by comparing the ratio of weighted sum of outputs to the weighted sum of inputs across DMUs.

Koopmans (1951) is associated with the first definition of technical efficiency: a DMU is technically efficient if and only if an increase in one of the outputs cannot be made without either decreasing some of the other outputs, or increasing the inputs. Debreu (1951) proposed the first radial measure of productive efficiency: a coefficient of resource utilisation. Farrell (1957) drew on the work of these two economists and proposed a method to empirically calculate the relative efficiency of several DMUs. He also suggested a way to decompose productive efficiency into two measures: technical efficiency and allocative or price efficiency.

Farrell (1957) is considered to be the first author to lay down the foundations of DEA, by proposing the construction of an empirical production frontier from several observed levels of input and output. This defined inefficiency as the radial distance from a point below the frontier to a point on the frontier. This distance represents the maximal radial expansion of the outputs while keeping inputs constant (output orientation), or the maximal radial contraction on the inputs, while keeping outputs constant (input orientation). One minus the equi-proportionate reduction in all inputs is a measure of input oriented radial efficiency. A score of unity is a necessary condition for relative efficiency. Similarly, a score of less than unity means that there is at least another DMU achieving the same level of outputs, using *fewer* inputs.

Charnes, Cooper and Rhodes (1978) extended Farrell's work and defined a linear program in order to obtain efficiency values, for a production system characterised by multiple inputs and multiple outputs. They also introduced an innovation to Farrell's measure of efficiency. They introduced slack measures to account for non-radial adjustments to the frontier. This model became known as the first DEA model, the CCR Model. We now present a graphical illustration of the measurement of efficiency by using DEA.

2.2.1 Graphical representation of DEA – radial and non-radial efficiency measures

Let us assume that there are 6 DMUs (A, B, C, D, E and F) using two inputs (x_1 and x_2) to produce one output (y). For simplification, let us also assume that all six DMUs produce one unit of output. Figure 2.1 below represents the input space of the production technology. Units A, B, and C are technically efficient because they are located on the frontier of technical efficiency. These units envelop all the other units, which are located to their right. The space to the right of the convex frontier is known as the *Production Possibility Set*. We have no evidence that it is possible to produce one unit of output y , using a combination of inputs situated to the left of the convex frontier formed by points A, B, and C. The vertical extension from point A upwards is also part of the frontier, but is a dominated part of the frontier, given that any point along this vertical line, is dominated by point A. Here, there is evidence that one unit of output y can be produced using the same quantity of x_1 and less quantity of x_2 . Any point along this vertical line is said to be *weakly efficient*, and a slack quantity of input x_2 can be identified. Reduction of the input usage by that amount of slack would make the DMU technically efficient. A similar situation happens to any points along the horizontal line from point C.

Units D, E and F are technically inefficient, because it is possible to reduce both inputs, while producing the same amount of output. For example, in order for DMU F to become efficient, it has to reduce its inputs from point F to point F_T . This reduction in inputs is a *radial* reduction, without altering the input mix (that is the ratio between x_1 and x_2). Its technical efficiency (TE_F) can be calculated as:

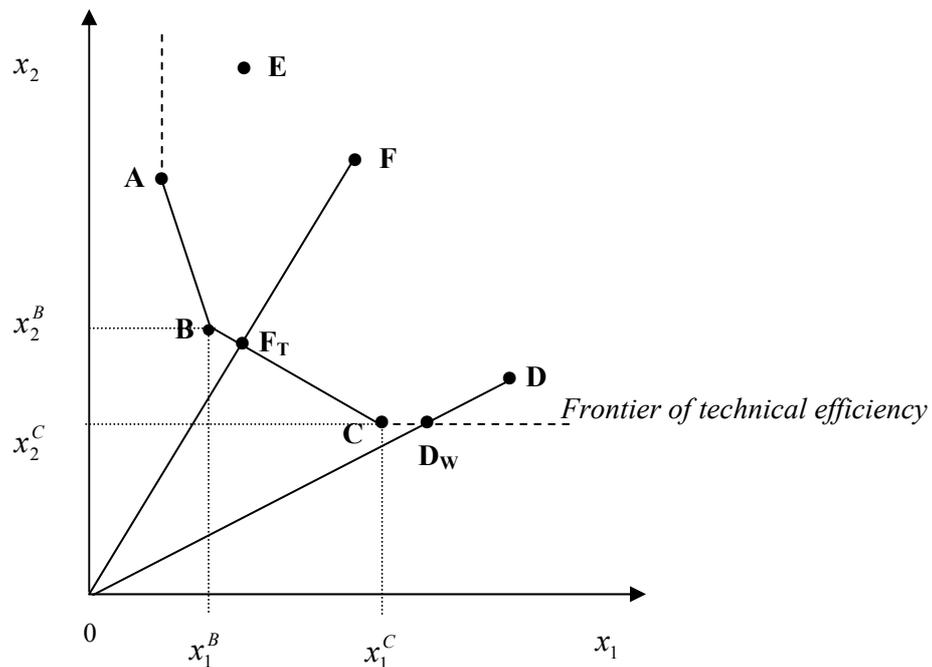
$$TE_F(x, y) = \frac{OF_T}{OF} \quad (2.1)$$

The optimal quantities of inputs for DMU F result from a convex combination of observed quantities of inputs used by DMUs B and C, by using the following formula:

$$F_T = \lambda_B(x_1^B, x_2^B) + \lambda_C(x_1^C, x_2^C) \quad (2.2)$$

where λ_B and λ_C are the intensity variables associated with each of the two peer DMUs. The projection of DMU F into F_T represents the shortest projection to the frontier, given that it respects the current mix of inputs used by DMU F. Assuming that DMU F wants to keep its current input mix, its peer group is constituted by DMUs B and C, with greater intensity being given to DMU B. DMUs B and C can provide DMU F with useful information on how to produce the same level of outputs using fewer resources, and the path F to F_T can be seen as a path for performance improvement over time. Using this path, several intermediate targets can be identified and progress towards the frontier can be monitored.

Figure 2.1: The measurement of technical efficiency



DMU D is a special case, because it is enveloped by the extension from the best practice frontier. DMU D is said to be *improperly enveloped*. After a radial reduction of all its inputs up to point D_w it is still possible to reduce one of its inputs without worsening the other input and without worsening the output level. In this example, a radial reduction of all its inputs to point D_w puts it on a position of weak efficiency or Farrell efficiency. In order to obtain a position of Koopmans efficiency, a further reduction of input x_1 is necessary to achieve point C on the frontier, which is a non-dominated point.

2.2.2 The measurement of technical efficiency – The CCR Model

Let us consider a general case characterised by the existence of n Decision Making Units ($DMU_j, j = 1, \dots, n$) to be compared, which use m inputs ($x_i, i = 1, \dots, m$) to produce s outputs ($y_r, r = 1, \dots, s$). We define the intensity variables attached to DMU_j as λ_j . The input weight or multiplier associated with input i is v_i and the output weight or multiplier associated with output r is u_r . Let us assume that constant returns to scale (CRS) prevail at the frontier, i.e., that changes in the inputs should lead to proportional changes in the outputs. Furthermore, let us take an input orientation to the frontier, i.e., we aim to minimise all the inputs, while keeping the outputs constant.

In order to simplify the notation used, let us define the following matrixes and vectors:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ y_{s1} & y_{s2} & \dots & y_{sn} \end{bmatrix}$$

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \end{bmatrix} \quad x_0 = \begin{bmatrix} x_{10} \\ x_{20} \\ \dots \\ x_{m0} \end{bmatrix} \quad y_0 = \begin{bmatrix} y_{10} \\ y_{20} \\ \dots \\ y_{s0} \end{bmatrix}$$

$$v = [v_1 \quad v_2 \quad \dots \quad v_m] \quad u = [u_1 \quad u_2 \quad \dots \quad u_s]$$

We can calculate the technical efficiency (TE) of DMU_θ by solving one of the linear programs presented in Model 2.1. The envelopment form and the multiplier form are two ways of presenting the same problem, given that one is the dual of the other.

Model 2.1: The CCR Model – input oriented

Envelopment Form	Multiplier Form
$TE_\theta(x, y)^{CRS} = \text{Min } \theta$ Subject to: $\theta x_0 - X\lambda \geq 0 \quad Y\lambda \geq y_0 \quad \lambda \geq 0.$	$TE_\theta(x, y)^{CRS} = \text{Max } uy_0$ Subject to: $vx_0 = 1$ $-vX + uY \leq 0$ $v \geq 0, u \geq 0.$

At this point we believe that it is prudent to define Farrell efficiency. A DMU is radially efficient if, and only if, it is not possible to improve all the inputs (or all the outputs) by a certain rate without worsening some other inputs or outputs. In this respect, a DMU is radial or Farrell efficient if the solution to Model 2.1 produces $\theta^* = 1$.

However, as previously shown, further non-radial input reductions or non-radial output augmentations may still be feasible. In order to identify these input shortfalls and output augmentations, we can solve a second stage DEA model, aiming at maximising the input and output slacks, while keeping $\theta = \theta^*$. In this respect, let us define the following ‘slack vectors’:

$$s^- = \theta x_0 - X\lambda, \quad s^+ = Y\lambda - y_0 \tag{2.3}$$

Also, let us define a vector of ones:

$$e = [1 \ 1 \ \dots \ 1], \text{ so that: } es^- = \sum_{i=1}^m s_i^- \text{ and } es^+ = \sum_{r=1}^s s_r^+ \quad (2.4)$$

In order to identify the maximum non-radial reduction in inputs (s^-) and the maximum non-radial improvement in outputs (s^+), we use the optimal value of θ^* determined from Model 2.1, and solve a second stage model (Model 2.2), using λ, s^- and s^+ as the variables.

We can now return to the definition of Koopmans efficiency. A DMU is Koopmans efficient if, and only if, it is not possible to improve any input or output without worsening some other input or output. That can be obtained by stipulating the following conditions in relation to the solutions of models 2.1 and 2.2:

$$\theta^* = 1 \text{ and } s_i^* = 0 \text{ (for all } i, i = 1, \dots, m) \quad (2.5)$$

$$\text{and } s_r^* = 0 \text{ (for all } r, r = 1, \dots, s) \quad (2.6)$$

Model 2.2: The CCR Model – Stage 2: maximisation of slacks

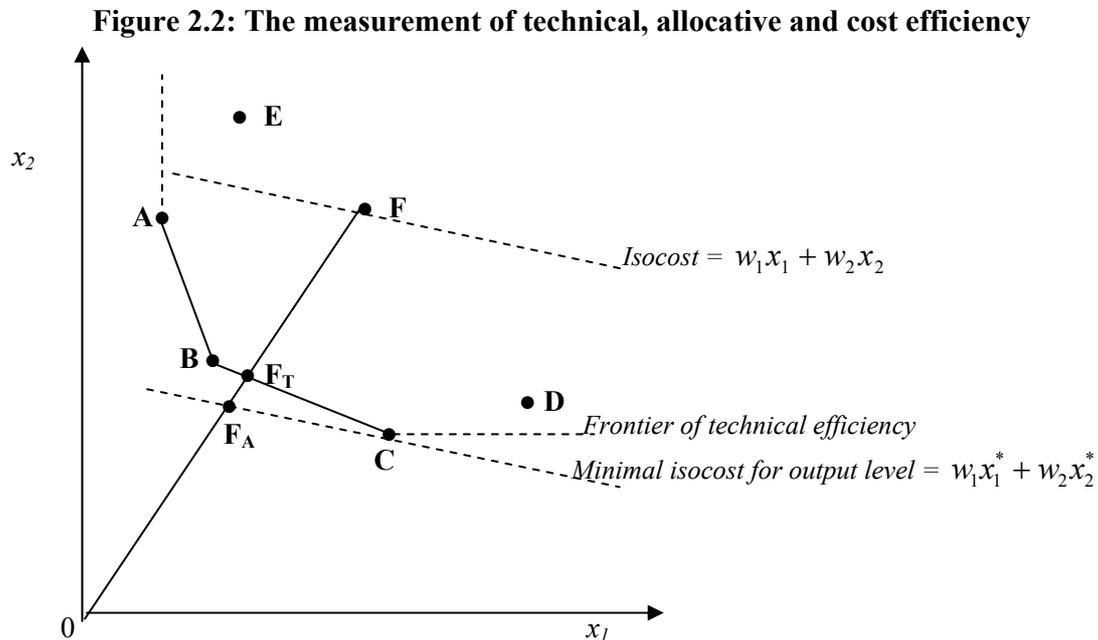
$$\begin{aligned} \text{Max } \omega &= es^- + es^+ \\ s^- &= \theta^* x_0 - X\lambda \\ \text{Subject to: } s^+ &= Y\lambda - y_0 \\ \lambda &\geq 0, s^- \geq 0, s^+ \geq 0. \end{aligned}$$

2.2.3 Measurement of allocative efficiency and cost efficiency When information about input prices exists, a DEA model can be formulated to measure cost efficiency, i.e., the efficiency of a DMU in producing a set of outputs at minimum cost. Figure 2.2 represents 6 DMUs (A, B, C, D, E and F) using two inputs (x_1 and x_2) to produce one unit of output (y). For simplification purposes, let us assume that all DMUs produce one unit of output. We have represented the frontier of technical efficiency and two isocost functions, given the current input prices (w_1 and w_2).

Based on this example, DMU C is the only unit that is cost efficient. Unit C is the only unit that produces the one unit of output at minimal cost: Point C is the point of intersection between the performance frontier and the minimal isocost. DMU C is said to be both technically and allocatively efficient. It is technically efficient because there is no evidence that it is possible to reduce all the inputs whilst producing one unit of output. It is allocatively efficient because it uses the optimal input mix given the current input prices. Units A and B are only technically efficient, given that their current mix of inputs is sub-optimal. When we consider the current input prices, both DMUs A and B use too much of input x_2 , when in fact they should use greater quantities of input x_1 . If these two DMUs operated with the same mix as that of DMU C, they could produce the same amount of output with reduced costs. DMU F is both

technically and allocatively inefficient. Its measure of allocative efficiency is calculated as:

$$AE_F(x, y) = \frac{OF_A}{OF_T}. \quad (2.7)$$



Cost efficiency can be obtained by multiplying the measure of technical efficiency by the measure of allocative efficiency (Farrell 1957):

$$OE_F(x, y) = \frac{OF_T}{OF} \times \frac{OF_A}{OF_T} = \frac{OF_A}{OF}. \quad (2.8)$$

A DEA model can be formulated to measure allocative efficiency. Let us define a price vector containing the unit prices of each input: $w = [w_1 \ w_2 \ \dots \ w_m]$. Assuming constant returns to scale and taking an input orientation, the minimum cost (MC) of production for DMU_0 can be calculated by solving the Linear Program presented in Model 2.3 (Färe, Grosskopf and Lovell 1994).

Model 2.3: The cost minimisation model

$$MC_0(x, y, w)^{CRS} = \text{Min } wx_0$$

Subject to:

$$x_0 - X\lambda \geq 0$$

$$Y\lambda \geq y_0$$

$$\lambda \geq 0.$$

Model 2.3 aims to minimise the total input cost, allowing the input mix to be changed, in a way that suits the current input prices. This is in contrast with the envelopment form of model 2.1 presented above, where the input mix is assumed to

remain constant, and only a radial reduction is allowed. This difference is evident in the first set of restrictions, where in Model 2.3 the quantities of inputs are variables to be optimised. Once the optimum quantities of inputs have been identified, the minimum feasible cost for DMU₀ can be calculated and is equal to: $MC_0(x, y, w)$. This value represents the minimum feasible cost, at which one unit of output for DMU₀ can be produced, considering the current input prices. If DMU₀ is cost efficient, then its current cost (w_0x_0) is already the minimum cost, and its cost efficiency is equal to one. Otherwise, its current cost is superior to the minimum feasible cost, and its cost efficiency is smaller than one. In this case, the vector of optimal input quantities is given by

$$x_0^* = [x_{10}^* \quad x_{20}^* \quad \dots \quad x_{m0}^*]^T \quad (2.9)$$

where the superscript T indicates vector transposition. Cost efficiency (also known as overall efficiency) can be calculated using the following formula (Färe, Grosskopf and Lovell 1994):

$$\text{Overall or cost efficiency} = OE_0(x, y, w) = \frac{MC_0(x, y, w)}{w_0x_0} \quad (2.10)$$

Cost efficiency includes the two components of efficiency. Firstly, it includes the component of technical efficiency, which does not take into consideration the input prices: it focuses solely on input quantities, and aims at achieving the maximum possible radial reduction. Secondly, it includes an allocative component, which takes prices into consideration and aims to choose the optimal input mix, given current input prices. If we want to isolate this allocative component, we need to run both Model 2.1 and 2.3, and calculate the allocative component using the following formula (Färe, Grosskopf and Lovell 1994):

$$\text{Allocative efficiency} = AE_0(x, y, w) = \frac{OE_0(x, y, w)}{TE_0(x, y)} \quad (2.11)$$

Given this relationship, cost efficiency can be decomposed into two components as shown below:

$$OE_0(x, y, w) = AE_0(x, y, w) \times TE_0(x, y) \quad (2.12)$$

So far we have focused on the measurement of technical and allocative input efficiency. The measurement of cost efficiency assumes the existence of input prices. If information regarding the exact input prices is not available, an alternative is to include information regarding a range of technologically realistic trade-offs between the different inputs. Similarly, information regarding the technologically realistic trade-offs between the different outputs can also be included in order to obtain more reliable estimates of efficiency. In DEA, this type of information is introduced via the use of weight restrictions. We will now briefly review the studies that have applied DEA to primary health care.

2. Previous studies: DEA applied to primary health care

The number of applications of DEA to primary health care is limited (Hollingsworth, Dawson and Maniadakis 1999). A substantial part of these studies have used DEA to compare primary health care providers in the UK (Szczepura, *et al.* 1993, Thanassoulis, Boussofiane and Dyson 1995, Bates, Baines and Whyne 1996, Salinas-Jiménez and Smith 1996, Giuffrida 1999, Buck 2000 and Giuffrida and Gravelle 2001). Pina and Torres (1992), Göni (1999) and Garcia *et al.* (1999) compared the efficiency of Spanish primary care centres. Luoma *et al.* (1996) used DEA to compare the efficiency of Finish health centres, and Zavras *et al.* (2002) have used DEA to compare Greek primary care centres. In America, DEA has been used to compare primary care clinics (Huang and McLaughlin 1989) and to compare mental health care programs (Schinnar *et al.* 1990). DEA has also been applied to Health Maintenance Organizations (HMOs) (Rosenman, Siddarthan and Ahern 1997, Siddarthan, Ahern and Rosenman 2000), and individual physicians (Chilingerian and Sherman 1996, 1997, Ozcan 1998, Wagner, Shimshak and Novak 2003).

A large number of applications of DEA have focused on the measurement of efficiency of hospitals. Examples include Nunamaker (1983), Grosskopf and Valdmanis (1987), Rosko (1990), Ozcan (1992), Byrnes and Valdmanis (1994), Parkin and Hollingsworth (1997), Mobley and Magnussen (1998), Solà and Prior (2001) and Tsai and Molinero (2002). A full review of these studies is outside the scope of this paper, as our main focus is on primary care evaluation.

Given the wide range of services delivered by primary care providers, the identification of outputs of primary care is a great challenge. A common feature in most of the studies reviewed here is that the number of patients registered is used to measure output. The use of the number of patients registered is particularly common in America, due to the publication of this data by HMOs. For example, Ozcan (1998) used the number of patients registered as outputs, separated by severity categories: low severity, medium severity and high severity. Chilingerian and Sherman (1996 and 1997) also used the number of patients registered with each physician, in this case separated into sex and age groups.

There are two problems associated with using the number of registered patients to compare the performance of primary health care providers. Firstly, it assumes that all the patients registered with a particular health care organisation are receiving the necessary services. Secondly, it assumes that the services provided to the patients registered are appropriate and of similar quality. These problems have been addressed by some of the studies and we shall focus on them.

Recognising that having patients on the lists does not necessarily mean service delivery, we find some studies that have used the number of consultations undertaken as an output measure in primary health care. Huang and McLaughlin (1989) used the number of consultations by physicians, nurses and non-health care professionals. Pina and Torres (1992) used output measures related to the average number of consultations per inhabitant and the average number of consultations by professional. Similarly, Luoma *et al.* (1996) included several output measures related to the number of patient visits to physicians.

Using the Whole Time Equivalent (WTE) health care professionals as inputs and using the number of consultations as outputs will lead to a system that rewards

shorter consultations. For this reason, Göni (1999) has used output measures related to the number of consultations and also related to the time devoted to the consultations. However, despite the fact that a consultation is an opportunity for contact with the health care professional, frequent consultations, per se, do not mean *better* primary health care. It is desirable to include measures of output that aim to capture the appropriateness of the consultations in terms of the delivery of the services deemed necessary to improve health outcomes. We will now review some studies that have attempted to include data regarding the appropriateness of service delivery.

Szczepura *et al.* (1993) studied the efficiency and effectiveness of General Practices in the region of Warwickshire in England. Unfortunately, unavailability of adequate data prevented the authors from building a more comprehensive model. The outputs used to measure efficiency refer solely to preventive care.

Schinnar *et al.* (1990) have also investigated the relationship between the efficiency and effectiveness in their comparison of 54 mental health care programs in the USA. They concluded that for lower than average levels of efficiency, there is a positive correlation between efficiency and effectiveness. However, for higher than average levels of efficiency, they were unable to find a clear relationship. Moreover, it was suggested that at an average level of efficiency, programs may be able to achieve the highest levels of effectiveness.

Thanassoulis, Boussofiene and Dyson (1995) used data on the provision of perinatal care in England to build three alternative DEA models for target setting. The main objective of the study was the proposal of several technical enhancements in the DEA model formulation.

Chilingerian and Sherman (1996, 1997) were also concerned with incorporating managerial preferences for target setting with DEA. They used a Cone Ratio DEA model to incorporate this information. Both of the studies focus on the same aspect: the evaluation of practice patterns of individual physicians, in order to achieve resource savings. The model also provides targets for performance improvement according to the chosen practice pattern. However, it was suggested that more research is needed to improve the measures of service quality, before we can be confident that these results genuinely represent resource saving potential.

Salinas-Jiménez and Smith (1996) were also concerned with the need to include quality indicators in the measurement of efficiency. In this study, data from 90 District Health Authorities in England is used to compare their relative efficiency. Due to non-availability of outcomes data, the DEA model constructed uses structure and process indicators as proxies for the quality of the service, such as the proportion of practices employing a practice nurse and the proportion of practice premises that satisfied the minimum quality standards. Salinas-Jiménez and Smith (1996) recognized the potential of DEA as a benchmarking technique in primary care, whilst emphasizing the need to include more reliable measures regarding the quality of services delivery in this context.

García *et al.* (1999) made another attempt to include information about the quality of primary care services in a DEA study. They studied the efficiency of 54 Primary Health Centres in the province of Zaragoza, Spain. Using the results from a previous study analyzing the opinion of primary health care managers in Spain (Urbina *et al.* 1997), some distinctive output variables were identified. From a managerial

perspective, two of the most important output variables were found to be: (1) ‘The Synthetic Index Product’, which provides “information about the coverage of the different services and the special effort (burden) that the services imply” (García *et al.* 1999: 71); and (2) ‘The Minimal Technical Standard’, which measures the compliance with the minimum standards of quality in the delivery of preventive and curative services. These two variables were used in this study, respectively, as a measure of the access to services according to need, and the quality of services delivered. Despite the progress made in the output variables used, the authors still pointed to the need to improve the definition of quality and case-mix in primary care in order to undertake a more comprehensive evaluation of providers.

Our review of the literature so far points towards the fact that most DEA studies applied to primary care focused on structure and outputs, but have not included outcomes. It is argued however that evaluation in general practice does need to be holistic in the sense that it is the general service as a whole that should be evaluated rather than the individual aspects and in principle DEA can provide such a methodology.

Evaluation in public services tends to be of a summative nature involving rankings, classifications and targets often set arbitrarily and in isolation from the overall performance. This approach can be dysfunctional and it is argued that for service improvement a formative approach, engaging the decision-makers is appropriate. It is thus argued that for DEA to be useful in performance improvement in primary care it needs to be deployed in the formative mode.

In summary it is argued that evaluation and performance improvement in primary care needs to be inclusive of structure, process, outputs and outcomes; that evaluation in primary care must be holistic in recognition that it is a general health care service; that evaluation must be formative, engaging stakeholders, if it is to lead to performance improvement; and in principle DEA provides the basis of an appropriate methodology. In the following section a model is developed for holistic, inclusive and formative evaluation of primary care.

4. Development of a new conceptual framework to compare primary care providers

In this section we discuss the development of a new conceptual framework for formative evaluation of primary care providers. This framework was developed during a series of workshops with Board members of two PCGs and two PCTs in England. This conceptual framework is to be used at the level of GP surgeries and aims to establish a link between local needs, resources usage, service delivery and outcomes achieved in primary care.

Our purpose of undertaking a performance assessment in primary care was a formative one, from the perspective of a PCG/T. In order to identify the criteria for evaluation and to decide on how to measure each criterion, we used the workshops with the participating PCG/Ts to develop and validate a simplified representation of primary care delivery in England.

4.1. A new conceptual framework to compare primary care providers

Figure 4.1 presents the new conceptual framework developed to compare primary care providers. In the rectangular light grey boxes we have the elements that characterise primary care delivery, the elements that drive it and the impact of this activity on the population's health. Graphically shown below these boxes are four criteria that can be used for performance assessment: (1) equity, (2) efficiency, (3) service effectiveness and (4) cost effectiveness. For example, efficiency is considered to be a measure of how well each provider is able to use its resources to deliver relevant services of high quality. An important aspect we would like to emphasize is that the measurement of the different dimensions of performance requires the use of different models. A single model would not be appropriate to measure all the dimensions as they relate to different parts of the process.

The simple 'one way' causality assumed to exist between needs, resources, service delivery and outcomes achievement represents a strong abstraction from the complex problem that characterises health care delivery. In order to make an attempt to conceptualise some of the complexities involved, we have added other elements to this simple representation.

In the rectangular dark grey boxes, we have factors that impact upon the performance of each surgery, but are not directly controllable by the surgery. For example, perceptions of local health and need will drive the demand for primary care services and surgeries have to respond to that demand by consulting the patients. The guidelines and targets imposed by the government will necessarily have an impact on the objectives and priorities followed by each surgery. The surgery's past achievement in terms of performance can also be expected to influence the level of resources allocated to a particular type of service delivery, particularly if past performance results were the subject of political scrutiny (Smith 1992). Finally, the characteristics of the population covered by each surgery, patients' propensity to consult and willingness to comply with treatment will also have a great influence on the level of resource usage, service delivery and outcomes achievement.

It is important to recognise that decision-making and resource allocation is not only driven by utility considerations. In fact, a number of multiple and potentially conflicting values inform decision-making from individual need considerations to personal gain, habit and competence.

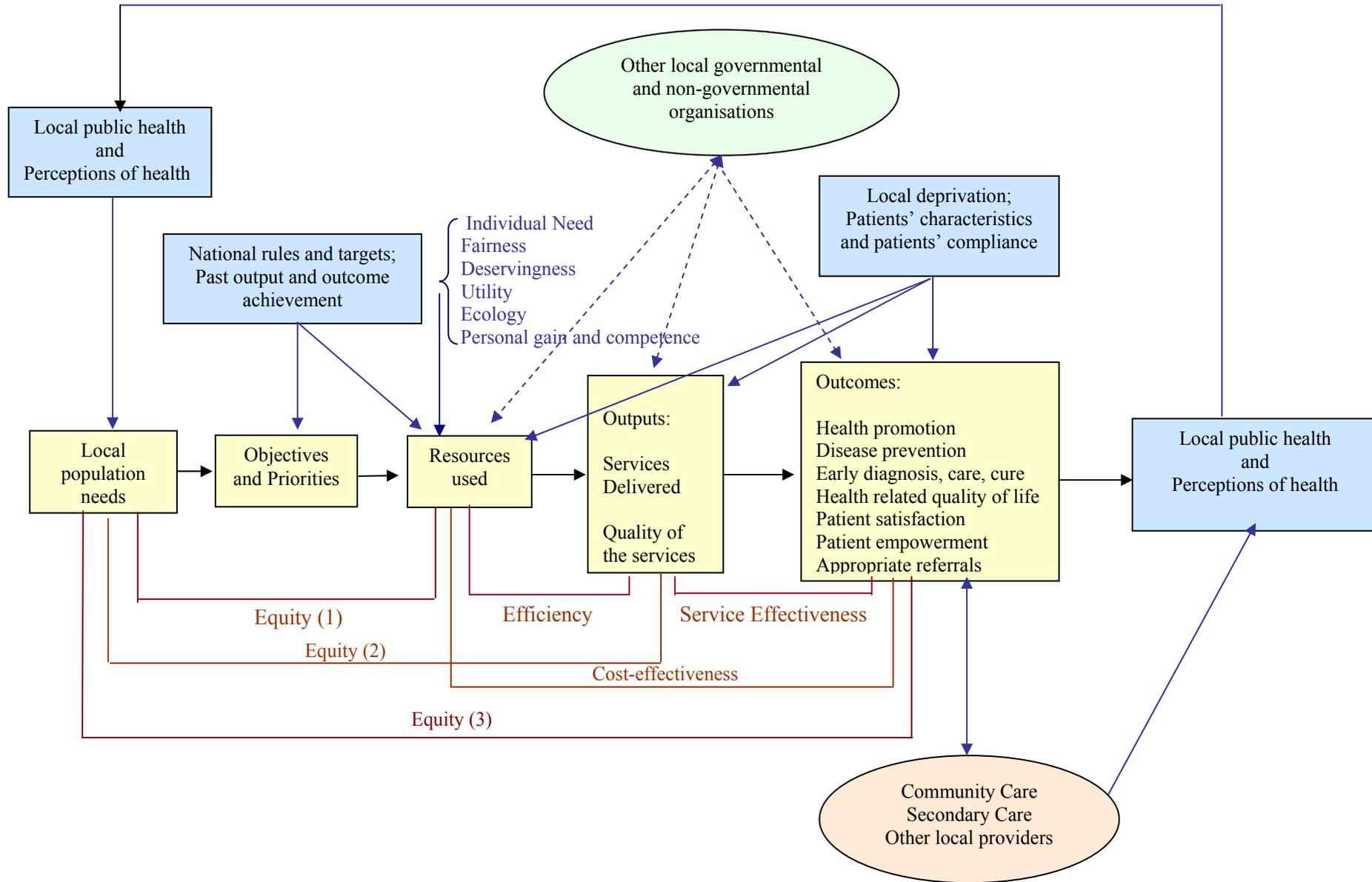
The elliptical boxes represent the organisations with which GP surgeries most directly engage, and which indirectly affect their performance. For example, the quality of the services provided by the local hospital can certainly be expected to impact on the level of resource usage, service delivery and outcomes achievement of the surgery.

This type of representation is not intended to accurately describe primary care delivery and the decision-making process associated with it. Instead, this representation is only intended to be an auxiliary device to understand how the different values and factors can relate to each other in a single decision. This leads to the recognition of the limitations involved in using a simple production metaphor to compare primary care providers.

Despite being a simplification, by focusing on the complexities involved in primary care delivery, this representation makes clear that a strong abstraction process is required in order to measure performance using the production metaphor. Firstly, if decision-making results from the heuristic application of multiple and potentially conflicting values, is it valid to assume that a clearly defined homogeneous ‘decision making unit’ (or agent) exists? We need to recognize that this metaphor results from an abstraction from many other aspects. However, we also need to recognize that such an abstraction can be useful, as it can lead to the identification of large discrepancies in the use of resources and the delivery of services, leading to a process of inquiry.

Finally, it is essential to recognise that performance assessment in the public sector will always be incomplete. Stewart and Walsh (1994: 47, emphasis added) make this point quite clearly. “The reason why adequate performance measures cannot be found in the public domain is, therefore, not a matter of technical problems alone, *it is inherent* in its character. It means that no set of indicators can ever be assumed to be complete, since in the public domain no relevant issues can be excluded”. In this respect, it is essential to recognise the limitations of any performance assessment framework in the public sector and proceed with great caution in terms of its use.

Figure 4.1: A conceptual framework for formative evaluation in primary care at the level of GP surgeries



4.2 Performance assessment criteria

During the workshops, four essential criteria were identified for performance assessment: (1) efficiency, (2) service effectiveness, (3) cost effectiveness and (4) equity. Below we discuss each one of these criteria and the methodology proposed to measure them.

4.2.1 The measurement of efficiency

One of the objectives of performance assessment in the public sector is to answer the question about whether an organisation is achieving 'value for money' (The Audit Commission 2000). This is, first of all, a managerial concern regarding the optimal use of public resources. In this respect, performance assessment is often associated with the measurement of efficiency.

The measurement of efficiency relies on the production metaphor. A set of inputs is used to produce a set of outputs. Inputs are the resources that are used to produce goods or deliver services. These may be materials or labour (human energy). Outputs are the goods produced or the services delivered. In primary health care, we are interested in services delivery. Services can be characterized as an action or set of actions that are desirable because of their effects (the outcomes). In particular, in health care, services delivery is desirable because it aims to help in maintaining or improving patients' health or patients' perceptions of it. In primary health care, this may be done by performing a diagnosis of patients' health status, by prescribing appropriate medication or treatment, or simply by providing advice or comfort to the patients.

There are two types of efficiency: (1) technical efficiency and (2) allocative efficiency. Technical efficiency assesses the relationships between the outputs produced and the resources used. It aims to maximize the amount of outputs produced from a certain level of resources usage (output orientation). Alternatively, it involves the minimization of resource usage in order to produce a certain level of outputs (input orientation). In order to perform a fair assessment of efficiency both the resources used and the outputs produced are assumed to be of similar quality.

As previously discussed, allocative efficiency relates to an assessment of the optimality of the mix of resources used given their current prices and the optimality of the outputs produced given the relative valuations placed on them. A provider may be technically efficient, but allocatively inefficient. If the perfect set of values could be identified for the resources and the outputs, a measure of overall efficiency could be obtained.

In the public sector, we are usually faced with situations where we do not have good measures for the prices of the resources or for the valuations of the outputs. Therefore, the use of a technique like DEA is advantageous, because it allows each provider to choose the best set of weights for its inputs and outputs. In this respect, and in the absence of any information regarding the trade-offs between the different inputs and outputs, DEA legitimizes all possible trade-offs and provides a conservative measure of technical efficiency.

Therefore, we express the measurement of technical efficiency using the following formula:

$$\text{Technical Efficiency} = \frac{\text{Weighted Sum of Outputs}}{\text{Weighted Sum of Inputs}} \quad (4.1)$$

This is a relative measure and for our study, it results from a comparison of several different providers.

4.2.2 Service effectiveness and cost effectiveness

Carter, Klein and Day (1992) discuss the ambiguity and confusion associated with the concept of effectiveness. An effective organisation can be defined as an organisation that achieves its objectives. Schinnar *et al.* (1990) proposed an alternative definition of effectiveness that we find very useful. They defined effectiveness as the ratio of outcomes achieved to services delivered and proposed the use of DEA to compare the effectiveness of different organisations. This is done by using an output-oriented DEA model, which aims to maximise the level of outcomes achieved, given the amount of services delivered; and aims to evaluate the ‘added value’ of the services delivered, independently of the resources used. Lovell, Walters and Wood (1994) have used DEA to measure service effectiveness in the education sector. Porto (2000) used this approach to compare the service effectiveness of substance abuse programs.

In this study, we use the definition proposed by Schinnar *et al.* (1990) to measure service effectiveness. Furthermore, we have extended this approach to measure cost effectiveness. This can be done by solving a DEA model that aims to minimise the total cost of production (or the level of resources used), whilst maintaining the amount of outcomes achieved. Alternatively, a comparison can be established by aiming to maximise the outcomes from a given level of resource usage. Below we present the formulas proposed to calculate service effectiveness and cost effectiveness:

$$\text{Service Effectiveness} = \frac{\text{Weighted Sum of Outcomes}}{\text{Weighted Sum of Outputs}} \quad (4.2)$$

$$\text{Cost Effectiveness} = \frac{\text{Weighted Sum of Outcomes}}{\text{Weighted Sum of Inputs}} \quad (4.3)$$

The formulation presented for the measurement of service effectiveness highlights the potential trade-offs between efficiency and service effectiveness given that in terms of efficiency measurement; the outputs appear in the numerator whilst in the measurement of service effectiveness, they appear in the denominator. The need to make this trade-off explicit was emphasised during the workshops. It was believed that one of the ways to improve service effectiveness was by focusing on the delivery of services towards a restricted number of more cooperative patients.

Despite the importance of ‘Value for Money’ assessments, this type of assessment is only concerned with the value of utility. Concerns for individual need

and fairness are also legitimate and performance assessment cannot ignore these values. As pointed by Ozcan and Smith (1999: 1), “it may be the case that perfectly efficient managerial solutions may be unacceptable to patients, and can be implemented only if they also satisfy additional criteria relating to fairness”. We now discuss the criteria of equity in terms of performance assessment and three alternative approaches for its measurement.

4.2.3 Equity between groups of patients

Equity is concerned with justice and fairness in the treatment of cases. With regards to health care provision, equity is concerned with justice in the treatment of different patients and particularly in the treatment of different groups of patients. An equitable health care system provides services of the same standard independently of, for example, patients’ income group, age, sex and area of residence.

Aristotle provided two essential definitions of equity. (1) Horizontal equity: equal treatment of equals; and (2) vertical equity: unequal but fair treatment of unequals. Each one of these definitions can be operationalised in a number of ways, and it is within this process that one finds divergences. For example, how do we assess ‘equality’? The most basic criterion is one of being human, and in that respect one can say that there should be equal treatment of human beings (Culyer and Wagstaff 1992). This criterion is behind the health care budget *allocation* of some countries, where the formulas for distribution aim at equal expenditure per capita. This approach fails to recognise an important aspect, that of *need*. Human beings can have different levels of health care need, and equality of treatment only becomes meaningful if there is need for that treatment. Therefore, horizontal equity can be re-defined as ‘equal treatment for equal need’. Identically, vertical equity would be re-defined as ‘more favourable treatment of those with higher need’.

Given the difficulties associated with the assessment of vertical equity, in this study we have concentrated on assessing horizontal equity only. We will now clarify what we mean by *equal need* and *equal treatment*.

4.2.3.1 The concept of need

Need for health care is an ambiguous concept, and we can find different definitions in the literature, as well as different approaches to its assessment. One definition that has been used in several recent studies (Smaje and Le Grand 1997, Waters 2000, Urbanos-Garrido 2001) equates need with patient self-reported illness. For example, Smaje and Le Grand (1997: 488), define need for health care as “the presence of long-term illness which limits activities and/ or an illness experienced within the previous two weeks”.

We believe that using a single indicator of need derived from patient reported illness has some limitations. Firstly, there are different levels of need, given that the health care needs of a patient reporting being ill in the previous week (for example having flu) cannot be compared with the health care needs of a patient who reports suffering from a chronic illness. Secondly, different groups of patients may have different pre-dispositions to recognize and report morbidity (Goddard and Smith

1998). Finally, and more fundamentally, equating self-reported illness with need for health care can be misleading for two reasons. (1) A patient might need health care without being ill, or perceiving to be ill. Preventive health care and disease screening are two examples. An undiagnosed person suffering from diabetes might feel perfectly fine at the moment, but is in fact in great need of health care. (2) A patient might be ill or perceive to be ill and not need health care, in the sense that currently available health care will not be of benefit to the patient.

Williams (1974) and Culyer (1976) proposed an early definition of need, which in our view, is preferable. This definition equates need for health care with the patient's capacity to benefit from the consumption of health care. In this respect, if someone cannot benefit from health care, although *ill*, he/ she is not said to be in need of health care. More recently, Culyer and Wagstaff (1992) proposed a modification to this definition in order to quantify need. They define need as the minimum amount of expenditure required to eliminate patients' capacity to benefit from health care. Despite the obvious advantages of such a definition, we believe that its operationalisation can bring additional difficulties. The amount of resources needed to eliminate the patients' capacity to benefit will depend on the efficiency of the health care system as well as on the effectiveness of the services provided. We have to assume that the term *minimum amount* of expenditure refers to the resources needed to deliver the necessary services, using efficient and effective systems. Given the difficulties involved in estimating this minimum amount of expenditure, we choose to work with the earlier definition where need is equated with capacity to benefit, although acknowledging that different patients can have different levels of need.

One question remains to be answered. How is the judgment about the capacity to benefit made? In practice, this judgment is made by health care professionals, patients and carers and it is the interplay of these judgments that drives health care delivery. In this study, we consider that a group of patients has health care need *if it is clinically expected* that their health status would improve (or deteriorate slower) due to receiving appropriate primary care. Having clarified what we mean by need for primary care, we now turn to the conceptualisation of equality of treatment.

4.2.3.2 The concept of equality of treatment

Equality of treatment can be measured using three criteria (Culyer 1991):

- (1) Service inputs - resources used in the provision of health care for population groups;
- (2) Service outputs - services delivered to population groups;
- (3) Service outcomes - the impact of the services delivered on population groups.

Based on these three criteria to assess equality of treatment, we can operationalise three definitions of horizontal equity:

- (1) Equal resources for equal need;
- (2) Equal service provision for equal need;
- (3) Equal health care outcomes for equal need.

The first definition (inputs/need) is an important one, and tends to be the one that is most used. However, it can be misleading, if we do not adjust for the relative

inefficiencies of the providers. For example, the fact that provider A has a higher ratio of expenditure for need than provider B, does not imply that the patients serviced by provider B are suffering from inequity of health care. In fact, they might receive a relatively higher level of services per need. This is because provider A might be inefficient when compared to provider B.

The second definition proposed (services/need) overcomes this problem, when we consider that it compares the amount of services provided with the needs of different population groups. At the most basic level, this definition is concerned with equality of access to services. However, given the difficulties in assessing equality of access, most studies have focused on equality of utilisation of health care. This poses some challenges because different people have different propensities to consult. Goddard and Smith (1998) state that it is important to distinguish between the inequity that is caused mostly by poorer supply of services to certain groups and the inequity that is explained mostly by different propensities to utilise services. These two types of inequity require very different types of policy intervention.

Definition (2) does not consider how effective each service provider is. It might be the case that provider A, although inefficient, delivers the *right* services to patients, in such a way that the effect these services have on the patients is higher than the impact of the services provided by B. If we are concerned with an equitable distribution of health care outcomes, then we need a different definition. The last definition (outcomes/need) is the most ambitious of all; it requires equal health outcomes across the different groups of patients. Regardless of health care efficiency and effectiveness, this definition requires that all groups have equal levels of health care outcomes for equal needs. In this respect, its achievement requires *positive discrimination*, towards those patients for which health services are less effective.

Despite the support shown by some researchers towards the use of definition (3) (Culyer and Wagstaff 1992), we feel that its use poses some challenges. We are back into the fact that people have different attitudes towards health care and towards health. The goal of 'outcomes equality' presupposes that everyone has the same preference for 'being healthy', puts the same effort into 'being healthy' and has the same attitude to risk. Mooney and Jan (1997: 82) make this point: "Even if we had a society which started from a position of equal health (with the same genetic endowment for all) and even if we all had perfect knowledge of the health effects of all activities in which we might indulge [and had equal access to effective health care], it is unlikely that within one year or two we would all have the same level of health". The question becomes whether it is reasonable to measure health care equity by assessing equality of outcomes. Sen (1992) states that equity is better assessed in terms of freedom to achieve rather than actual achievements. Others might propose that we should discard definition (3), and work solely with definitions (1) and (2), focusing particularly on equality of access. Fair access to services is in fact one of the founding principles of the NHS and the White Paper published in 1997 makes it one of the six dimensions of performance. Nevertheless, we feel that a comparison of health outcomes across different groups is relevant because it allows us to assess if different groups appear to be systematically subject to less effective care and point towards providers who may need extra resources in order to secure similar standards in terms of outcomes achievement.

4.2.4 Conflicts between the different performance criteria

One consequence of using several criteria for evaluation is that this may lead to alternative and potentially conflicting views of best practice. Furthermore, if the objective is to identify strategies for performance improvement, it may lead to several potentially conflicting views regarding the need and direction for change. For example, in our case study, we found a set of surgeries that are performing very well under the equity criterion, but these surgeries are not necessarily efficient or effective in their services delivery. Furthermore, we have identified surgeries that are clinically effective, but are not efficient. The question is, do these surgeries need to change the way they operate, and if so, what direction for improvement should they take? This is in fact a difficult question, and one which appropriate discussion is outside the scope of this paper. Nevertheless, some tentative suggestions regarding methods to handling this question are proposed below.

One method is to admit that each organisation will have to undergo a process of negotiation and decide based on its priorities and values if it wants to change, and if so, what is the direction of change it wants to take. The problem with this solution is that it can lead to the perpetuation of the values of the most powerful stakeholders. We propose two alternative strategies regarding performance assessment and improvement whenever we are in presence of multiple incommensurable and conflicting objectives. The first alternative uses the concept of robustness proposed by Dryzec (1983):

“Optimisation is a legitimate procedure only if we are in possession of a single theoretical perspective in which we have a high degree of confidence, and believe there is only a single possible context in which the effects of policy will be felt. Given, though, that the components of our system are unreliable, one should pursue *robustness* rather than optimisation. A robust policy alternative is one expected to perform tolerably well across the whole range of scenarios *given* any one of the pertinent theoretical perspectives.”

Dryzec (1983: 360)

This alternative assumes that all values and criteria are equally legitimate and desirable. Furthermore, it assumes that a hierarchy of importance cannot be established between values. Therefore, no limits are imposed on the types of trade-offs that are to be accepted between the different values. However, depending on the context, it might be more appropriate to use a ‘qualified robustness’. This would entail an identification of the levels of trade-offs that were to be deemed acceptable between specific values and evaluation criteria. For example, we might want to impose that no decrease in the level of equity is to be accepted, independently of the increases in efficiency or effectiveness that it may generate. Le Grand (1991) defends that equity should be the primary objective in public services provision. Only when this criterion has been satisfied, should we be concerned with an improvement in the level of efficiency. Alternatively, we may say that the level of equity should never fall below a particular level, independently of the benefits it may generate in the other measures. A further alternative is to say that a decrease in the level of equity is only accepted if it leads to increases in the other evaluation criteria of a certain magnitude.

Independently of the criterion used to handle the trade-offs between the multiple values that underlie public services delivery, the important aspect to recognise is that this is a political decision that should be explicitly made. The complexity and responsibility involved in making such a decision should not be overshadowed by rhetoric about delegation of decision-making power. It is a political decision, whose effects should lead to a clear chain of central accountability independently of local decision-making (Du Gay 2000).

4.3 The impact of environmental variables in the assessment of performance in primary care

There are some factors outside the clinicians' control that can impact on the relative performance of each surgery. As a result it is important that these factors not be overlooked. During the workshops we identified the following factors that can contribute to the level of resources used, services delivered and outcomes achieved by each surgery:

- (1) The case-mix of the patients registered with each surgery, in terms of age, gender, ethnicity and health status;
- (2) The level of socio-economic deprivation of the area covered by each surgery. Socio-economic deprivation aims to measure the extent to which one area is disadvantaged in terms of the essential socio-economic factors. The Department of the Environment, Transport and the Regions (DETR 2000), suggests six domains of area deprivation: (1) income, (2) employment, (3) health deprivation and disability, (4) education skills and training, (5) housing and (6) geographical access to essential services. Area deprivation can influence the patients' propensity to pursue health care and the patients' lifestyle;
- (3) The level of patient compliance with the treatment and recommendations of the clinicians;
- (4) The quality of the services delivered by other local providers;
- (5) Random variation in the data.

4.4 The appropriateness of the production metaphor for primary care delivery

The importance of measuring outcomes and linking the achievement of outcomes with the level of resource usage and the level of services delivered relates with the need to make an assessment regarding which structures and processes of care appear to lead to best value for money. However, it has been stated that measuring the outcomes of primary care delivery poses considerable challenges (Harris 1993, Salinas-Jeminéz and Smith 1996). These difficulties relate to several essential aspects. Firstly, different groups of people have different objectives for primary care and therefore give value to different outcomes. This emphasises the pluralistic nature of primary care delivery and implies that any outcome assessment in primary care is necessarily partial. Secondly, some outcomes of primary care are only visible after many years of intervention, for example the outcomes of preventive care. Thirdly, some outcomes are difficult to quantify because they relate to subjective aspects of quality. For example, it is very difficult to assess the appropriateness of diagnosis and

the appropriateness of the support given to patients. Furthermore, in the absence of well-determined models of primary care delivery it is very difficult to establish a link between resources usage, services delivery and outcomes achievement.

As pointed out by Williams (1996), these difficulties should not imply that the task is impossible or inappropriate. Firstly, it is important to make clear from which perspective outcome is being measured and then it is important to arrive at outcome measures that have been shown to relate strongly with the activities undertaken by health care professionals. Furthermore, patients' characteristics or other non-controllable factors that are thought to have an effect on the relationship between resources usage and outcomes achievement should be taken into account when comparing the different providers (Salinas-Jeminéz and Smith 1996). We will next discuss alternative measures of health outcomes that can be used to compare primary care providers.

4.4.1 Measures of health outcomes for primary care

There are several types of measures that can be used to assess the outcomes of primary care. The first type of health outcome measure is rooted in the biomedical model of health, which takes a disease-based view of ill health. According to this model, health is the absence of disease diagnosed by using medical criteria. An alternative approach to outcomes assessment takes the patient's subjective experience of health and aims to measure health outcomes based on patient reported symptoms and feelings (Jenkinson 1994).

From a biomedical point of view, examples of health outcomes are mortality-based measures such as death rates, condition-specific death-rates, infant mortality, maternal mortality, life expectancy, suicide rates and measures of avoidable deaths. The use of mortality based indicators is relatively unsatisfactory to assess primary care providers because they occur in relatively small numbers, cover only the effects of a small proportion of the services delivered in primary care and, in many cases, it is very difficult to establish a link between the primary care delivered and mortality rates. Indicators of morbidity and disability can also be used as outcome measures.

An alternative to the use of mortality and morbidity rates, still using the biomedical model of health, relates to the use of indicators of physical well being of the patients. In this case, indicators of blood pressure and cholesterol control can be used as measures of health outcomes. The justification to use these measures relates to the fact that, in general, the probability of developing health complications is reduced if the essential clinical indicators are kept under control. The use of these indicators usually requires clinical evidence in terms of determining the exact thresholds of abnormality. The evidence regarding the optimal levels of clinical outcomes tends to be linked with the management of certain specialities, such as diabetes, heart disease and cancer (Hunter and Fairfield 1997).

The subjective measures propose an alternative to the traditional biomedical model of health outcomes assessment, by emphasizing the importance of patients' general well being and social functioning. Jenkinson (1994) identified three reasons to use subjective measures of health assessment. First, the difficulties involved in developing objective measures of outcome, with clear thresholds of abnormality. Second, the frequently reported discord between objective and the patients' subjective

assessments of health. Patients may feel ill without objective symptoms of disease and patients may feel very well whilst having a medical disease. Third, in cases of diseases with high levels of prevalence, patients who experience the disease may not see it as an illness. In this respect, it is suggested that the assessment of health outcomes should aim to capture both the objective and subjective accounts of health.

Subjective assessments of health are part of the broader concept of health related quality of life. Patrick and Erickson (1993: 22) define health related quality of life as “the value assigned to the duration of life as modified by the impairments, functional states, perceptions and social opportunities that are influenced by disease, injury, treatment, or policy”. Numerous instruments have been developed to measure health related quality of life in its numerous domains (Bowling 1991). These instruments are designed to measure the different domains of health related quality of life, including physical, cognitive and social functioning, self efficacy, well-being, life satisfaction, self-esteem and happiness (Albrecht 1994). Despite the importance of including patients’ subjective assessment of health care, there is disagreement between the appropriate domains that should be included in the assessment of health outcomes (Bowling 1997). This aspect is of fundamental importance if we are to fairly attribute the results of outcome assessment to the clinicians.

Health related quality of life measures have also been used to evaluate the cost-utility of different interventions by using Quality-Adjusted Life Years or QALYS (Williams 1985). This methodology aims to calculate a single indicator that estimates the duration of life weighted by the quality of life during that period. One year of healthy life is scored as one, but one year of unhealthy life is scored as less than one. The specific weights given to different states of health have to be elicited in an experimental setting with a sample of patients, or other stakeholders. The difficulty in eliciting these weights is one of the difficulties associated with this measure (McAlister 1994). Furthermore, it has been argued that outcome assessment should go beyond the measurement of the technical aspects of care in order to make an assessment regarding the quality of the interpersonal aspects of care, by using patient satisfaction surveys (McAlister 1994).

Patient satisfaction surveys aim to assess whether the access to the services is appropriate and the extent to which the services delivered are of acceptable quality. In this respect, satisfaction surveys provide information regarding the patients’ perspective about the performance of the primary care providers. Furthermore, satisfaction surveys can provide very important information regarding the most appropriate structures and processes of delivering services. The use of patient surveys in a performance assessment exercise can also represent a useful way of data triangulation. We do recognize the limitations associated with this method of health care evaluation. High levels of satisfaction tend to be reported, particularly if a general question about the level of satisfaction with the services is used. Furthermore, it has been suggested that patient satisfaction levels can be influenced by question wording and by the response format offered. For a discussion of these limitations please refer to Bowling (1997). Despite these limitations, we feel that the use of patient satisfaction surveys can be very valuable in providing an assessment from the patients’ perspective.

4.4.2 The need to focus on a specific disease

The feasibility of extracting data from primary care providers in the UK in order to implement the conceptual framework developed was discussed with the participating PCG/Ts. Our aim was to include both measures of objective and subjective health outcomes. We concluded that this exercise would be very difficult, given that all the data available for outcomes related to objective measures of specific diseases, such as diabetes, Coronary Heart Disease, cancer, asthma and hypertension. The measures available did not cover all the areas of activity of primary care and it was believed that it would be inappropriate to establish a link between resources usage, services delivery and outcomes achievement. Furthermore, it was believed that a formative evaluation focusing on a specific disease would provide more in-depth results regarding performance improvement.

The benefits of managing primary care by focusing on a speciality (also known as disease management) relate to the fact that this approach views the patients as experiencing the clinical course of a disease and uses the clinical evidence available to develop guidelines and protocols of care, aiming to improve the process and outcomes of care (Hunter and Fairfield 1997). The existence of clear protocols of care facilitates the establishment of a link between service delivery and outcomes achievement. Confidence in establishing this link is essential in order to compare providers based on a production metaphor (Brignall and Modell 2000).

In terms of our application of DEA to compare primary care providers, there are thus several reasons to focus on a speciality, such as diabetes care, rather than primary care as a whole:

- (1) the absence of appropriate outcome data for primary care as a whole;
- (2) the existence of relatively well defined measures of outcomes for diabetes care;
- (3) the absence of a 'well-determined' model of primary care delivery that works within the production metaphor;
- (4) the pluralistic nature of what good primary care means;
- (5) the existence of relatively well defined production functions (in terms of structure, process and outcomes) for diabetes care delivery.

In this respect, the fact that the 'product' characteristics of a particular speciality, such as diabetes care delivery, are more clearly defined than those of primary care as a whole better facilitates the construction of production models for performance assessment and makes the notion of formative evaluation possible. By focusing on a particular disease, the identification of needs for primary care is also facilitated.

However, we recognise the problems associated with the performance assessment of a single speciality in primary care delivery. Given its general practice focus, primary care professionals should aim towards a comprehensive and coordinated, patient-focused service delivery. A range of synergetic relationships can be explored between different specialities. Very often the preventive messages are similar across many specialities. Furthermore, the treatment of one health problem can interact with other health problems that the patient may have. Primary care

management by focusing on specialities has also been criticised because it may lead to the fragmentation of care for patients (Harris 1996), and also because it may lead to the potential loss of skills of primary care professionals in terms of the management of certain diseases (Bodenheimer *et al.* 1999). The importance of continuity of care is emphasised, together with the role of primary care physicians as coordinators of care. These issues suggest that efforts should be made to undertake an evaluation of primary care providers that recognises their role as coordinators of care. This should be the subject of future research.

5 Conclusion

In this paper we have presented a conceptual framework developed for formative evaluation of primary care providers. This framework aims to establish a link between the local needs, resources used, services delivered and outcomes achieved in primary care. By linking these four elements, four criteria for evaluation are proposed: (1) equity, (2) efficiency, (3) service effectiveness and (4) cost effectiveness. DEA is proposed as a methodology that can be used to measure efficiency, service effectiveness and cost effectiveness. Our measure of equity of services utilisation is the ratio between the services utilised and the explicit needs of the patients.

The difficulties encountered in the implementation of this framework to evaluate primary care as a whole are discussed. We conclude that the pluralistic nature of primary care delivery places difficulties in the measurement of outcomes for primary care. Furthermore, the fact that clear models of primary care delivery cannot be identified questions the appropriateness of the production metaphor for performance assessment at this level. The need to focus on a specialty is justified in order to better establish a link between resources usage, services delivery and outcomes achievement. Finally, the advantages and disadvantages of evaluating primary care providers by focusing on a specialty were discussed.

The study detailed in this paper makes the following contributions. Firstly, through the development of a new conceptual framework to compare primary care providers, we contribute to the under-researched area of performance assessment in primary care. Secondly, by developing a framework that links four different evaluation criteria, we contribute to a discussion regarding the potential conflicts between these criteria and ways to achieve an appropriate balance between them.

This study suffers from several limitations. Firstly, the use of a standard model of DEA does not take into account the influence of stochastic factors on performance. It can certainly be expected that part of the variability encountered in terms of efficiency and effectiveness is the result of natural stochastic factors. If a large sample of providers is available for comparison, we suggest the use of Stochastic Frontier Analysis (SFA) to separate between stochastic variability, the effects of environmental factors and pure inefficiency and ineffectiveness, according the methodology developed by Fried *et al.* (2002). Secondly, if data is available regarding the needs, inputs, outputs and outcomes for several periods, with confidence regarding the comparability of the data collection and extraction methods, we suggest the use of

Malmquist indexes in order to analyse the changes in productivity, efficiency and effectiveness in time. This is the subject of future research.

Future studies should also test the applicability of this framework to compare health care providers and investigate the relationships between the different evaluation criteria in different contexts, together with an analysis of the different factors that may explain why certain health care providers appear to perform better than others in certain criteria.

Acknowledgements

This study results from the research conducted towards my degree of PhD, under the supervision of Professor Robert G. Dyson and of Dr James E. Storbeck.

I would like to acknowledge financial support from the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia). I would also like to thank all my colleagues at Faculdade de Economia, Universidade do Algarve, for all their support throughout the period of study.

This study benefited greatly from meetings with the health care professionals from the participating Primary Care Groups and Trusts and GP surgeries. This study also benefited greatly from the comments and discussions during the seminars and other occasions with the Warwick DEA team: Robert Dyson, James Storbeck, Victor Podinovski, Estelle Shale, Joseph Coughlan and Susila Munisamy-Doraisamy. Furthermore, I would like to thank the two examiners of this PhD thesis, Professor Peter Smith and Dr Victor Podinovski. Their interest and constructive comments have greatly improved the quality of this work and have provided further motivation to submit it for publication. The responsibility for any errors is mine alone.

References

- Albrecht, G. L. (1994). 'Subjective health assessment'. In C. Jenkinson (eds) 'Measuring health and medical outcomes', London: UCL Press, 7-26.
- Bates, J., Baines, D. and Whynes, D. (1996). 'Measuring the Efficiency of Prescribing by General Practitioners', *Journal of the Operational Research Society* **47**: 1443-1451.
- Bryce, C., Engberg, J. and Wholey, D. (2000). 'Comparing the Agreement Among Alternative Models in Evaluating HMO Efficiency'. *Health Services Research* **35**: 509-528.
- Bodenheimer, T., Bernard, L. and Casalino, L. (1999). 'Primary Care Physicians Should Be Coordinators, Not Gatekeepers'. *Journal of the American Medical Association*, **281** (21): 2045-49.
- Bowling, A. (1997). 'Research Methods in Health: Investigating Health and Health Services'. Buckingham: Open University Press.
- Brignall, S. and Modell, S. (2000). 'An institutional perspective on performance measurement and management in the 'new public sector''. *Management Accounting Research*, **11**: 281-306.
- Buck, D. (2000). 'The Efficiency of the Community Dental Services in England: a data envelopment analysis'. *Community Dentistry and Oral Epidemiology* **28**: 274-280.
- Byrnes, P. and Valdmanis, V. (1994). 'Analysing Technical and Allocative Efficiency of Hospitals'. In A. Charnes, W. W. Cooper, A. Y. Lewin and L. M. Seiford (eds.). 'Data Envelopment Analysis: Theory, Method and Applications'. London: Kluwer Academic Publishers, 129-144.
- Carter, N., Klein, R. and Day, P. (1992). 'How Organisations Measure Success: The Use of Performance Indicators in Government'. London: Routledge.
- Charnes, A., Cooper, W. and Rhodes, E. (1978). 'Measuring the Efficiency of decision-making units'. *European Journal of Operational Research*, **2**: 429-444.
- Chilingerian, J. and Sherman, H. (1996). 'Benchmarking physician practice patterns with DEA: A multi-stage approach for cost containment'. *Annals of Operations Research* **67**: 83-116.
- Chilingerian, J. and Sherman, H. (1997). 'DEA and primary care physicians report cards: deriving preferred practice cones from managed care service concepts and operating strategies'. *Annals of Operations Research* **73**: 35-66.
- Cooper, W., Seiford, L. and Tone, K. (1999). 'Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software'. Kluwer Academic Publishers, Massachusetts, Kluwer Academic Publishers.
- Culyer, A. J. (1976). 'Need and the National Health Service: Economics and Social Choice'. London: Martin Robertson.
- Culyer, A. J. (1991). 'Health, health expenditures and equity'. Discussion Paper 83, York: Centre for Health Economics, University of York.
- Culyer, A. J. and Wagstaff, A. (1992). 'Need, Equity and Equality in Health and Health Care'. Discussion Paper 95, York: Centre for Health Economics, University of York.
- Debreu, G. (1951). 'The coefficient of resource allocation'. *Econometrica*, **19** (3): 273-292.
- Defelice, L. and Bradford, D. (1997). 'Relative inefficiency in production between solo and group practice physicians, *Health Economics* **6**: 455-65.

- Department of the Environment, Transport and the Regions (DETR) (2000). 'Indices of Deprivation 2000'; Regeneration Research Summary no. 31, London. Available online on <http://www.urban.odpm.gov.uk/research/summaries/03100/pdf/rrs03100.pdf>, accessed on 15/04/03.
- Dryzec, J. (1983). 'Don't Toss Coins in Garbage Cans: A Prologue to Policy Design'. *Journal of Public Policy*; **3**: 345-367.
- Du Gay, P. (2000). 'In praise of bureaucracy', London: Sage Publications.
- Dyson, R.G. (2000). 'Performance Measurement and Data Envelopment Analysis – rankings are rank!'. *OR Insight*, **13** (4): 3-8.
- Färe, R., Grosskopf, S. and Lovell, C. (1994). 'Production Frontiers', Cambridge: Cambridge University Press'.
- Farrell, M. (1957). 'The measurement of productive efficiency'. *Journal of the Royal Statistical Society A*, **120** (Part III): 253-290.
- Fried, H., Lovell, C., Schmidt, S. and Yaisawarng, S. (2002). 'Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis'. *Journal of Productivity Analysis*, **17**: 157-174.
- García, F., Marcuello, C., Serrano, D. and Urbina, O. (1999). 'Evaluation of Efficiency in Primary Health Care Centres: An application of data envelopment analysis'. *Financial Accountability and Management* **15** (1): 67-83.
- Giuffrida, A. (1999). 'Productivity and Efficiency changes in primary care: a Malmquist approach'. *Health Care Management Science* **2**: 11-26.
- Giuffrida, A., Gravelle, H. and Roland, M. (2000). 'Performance indicators for managing primary care: the confounding problem'. In P.C. Smith (eds). 'Reforming markets in health care: an economic perspective'. Buckingham: Open University Press, pp. 162-186.
- Giuffrida, A., Gravelle, H. and Sutton, M. (2000). 'Efficiency and administrative costs in primary care'. *Journal of Health Economics*, **19**(6): 983-1006.
- Giuffrida, A. and Gravelle, H. (2001). 'Measuring performance in primary care: econometric analysis and DEA'. *Applied Economics* **33** 163-175.
- Goddard, M., Mannion, R. and Smith, P.C. (2000). 'The performance framework: taking account of economic behaviour'. In P.C. Smith (eds). 'Reforming markets in health care: An economic perspective', Buckingham: Open University Press, 138-161.
- Goddard, M. and Smith, P. C. (1998). 'Equity of Access to Health Care'. Report, March 1998, Centre for Health Economics, University of York.
- Goddard, M., Mannion, R. and Smith, P. C. (1999). 'Assessing the performance of NHS Hospital Trusts: the role of 'hard' and 'soft' information'. *Health Policy*, **48**: 119-134.
- Göni, S. (1999). 'An analysis of the effectiveness of Spanish primary health care teams'. *Health Policy*, **48**: 107-117.
- Grosskopf, S. and Valdmanis, V. (1987). 'Measuring hospital performance: A non-parametric approach'. *Journal of Health Economics*: **6**: 89-108.
- Ham, C. (1999). 'Health Policy in Britain'. London: Macmillan Press Ltd.
- Harris, A. (1993). 'Developing a research and development strategy for primary care'. *British Medical Journal*, **306**: 189-1992.
- Harris, J. M. (1996). 'Disease Management: New Wine in New Bottles?'. *Annals of Internal Medicine*, **124** (9): 838-842.

- Hollingsworth, B., Dawson, P.J. and Maniadakis, N. (1999). 'Efficiency measurement of health care: a review of non-parametric methods and applications'. *Health Care Management Science* **2**: 162-172.
- Huang, Y-G 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.
- Hunter, DJ and Fairfield, G. (1997). 'Managed care: Disease management'. *British Medical Journal*, **315**: 50-53.
- Koopmans, T. (1951). 'An Analysis of Production as an Efficient Combination of Activities', in T. Koopmans (Ed.), 'Activity Analysis of Production and Allocation'. Cowles Commission for Research in Economics, Monograph 13, New York: Wiley.
- Jenkinson, C. (1994). 'Measuring health and medical outcomes: an overview'. In C. Jenkinson (eds) 'Measuring health and medical outcomes', London: UCL Press., 1-6.
- Le Grand, J. (1978). 'The Distribution of Public Expenditure: The Case of Health Care'. *Economica*, **45**: 125-42.
- Le Grand, J. (1991). 'The distribution of health care revisited'. *Journal of Health Economics*, **10**: 239-45.
- Lovell, C. Walters, L and Wood, L. (1994). 'Stratified models of education production using modified DEA and regression analysis' in Charnes, A., Cooper, W., Lewin, A. and Seiford, L. (Ed.). 'Data Envelopment Analysis: Theory, Methodology and Applications'. London: Kluwer Academic Publishers, 329-351.
- Luoma, K., Järviö, M-L., Suoniemi, I. and Hjerppe, R. (1996). 'Financial Incentives and Productive Efficiency in Finnish Health Centres'. *Health Economics* **5**: 435-445.
- McAlister. D. (1994). 'Putting Health Economics into Quality'. *Public Money and Management*, April-June: 15-22.
- Mobley, L. R. and Magnussen, J. (1998). 'An international comparison of hospital efficiency: does institutional environment matter?'. *Applied Economics*, **30**: 1089-1100.
- Money, G. and Jan, S. (1997). 'Vertical Equity: weighting outcomes? Or establishing procedures?'. *Health Policy*, **39**: 70-87.
- National Health Service Executive (1999). 'The NHS Performance Assessment Framework'. London: NHS Executive.
- 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** (2): 183-205.
- Ozcan, Y. A. (1992). 'Sensitivity Analysis of Hospital Efficiency Under Alternative Output/Input and Peer Groups: A Review'. *Knowledge and policy: The International Journal of Knowledge Transfer and Utilization* **5**: 1-29.
- Ozcan, Y. A. (1998). 'Physician benchmarking: measuring variation in practice behavior in treatment of otitis media'. *Health Care Management Science*, **1**: 5-17.
- Ozcan, Y. and Smith, P. C. (1999), 'Towards a science of the management of health care', *Health Care Management Science*, **1**: 1-4.
- Parkin, D. and Hollingsworth, B. (1997). 'Measuring production efficiency of acute hospitals in Scotland, 1991-94: validity issues in data envelopment analysis'. *Applied Economics*, **29**: 1425-1433.
- Patrick, D.L. and Erickson, P.E. (1993). 'Health status and health policy: allocating resources to health care. Oxford: Oxford University Press.

- Pedraja-Chaparro, F., Salinas-Jiménez, J., Smith, P. C. (1997). 'On the Role of Weight Restrictions in Data Envelopment Analysis'. *Journal of Productivity Analysis*, **8**: 215-230.
- 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**: 213-225.
- Porto, J. V. (2000). 'Comparing the Efficiency and Effectiveness of Substance Abuse Treatment Programs Using a Two-stage Data Envelopment Analysis'. PhD thesis.
- Rosenman, R., Siddharthan, k. and Ahern, M. (1997). 'Output Efficiency of health maintenance organizations in Florida'. *Health Economics*, **6**: 295-302.
- Rosko, M. (1990). 'Measuring Technical Efficiency in Health Care Organizations'. *Journal of Medical Systems*, **14**: 307-322.
- Salinas-Jiménez, J. and Smith, P.C. (1996). 'Data envelopment analysis applied to quality in primary health care'. *Annals of Operations Research*, **67**: 141-161.
- Schinnar, A. P. (1993). 'Tradeoffs between Efficiency and Effectiveness in Management of Public Services'. In Y. Ijiri (eds). 'Creative and Innovative Approaches to the Science of Management'. London: Quorum Books, 177-190.
- Schinnar, A. P., Kamis-Gould, E., Delucia, N. and Rothbard, A. B. (1990). 'Organizational Determinants of Efficiency and Effectiveness in Mental Health Partial Care Programs'. *Health Services Research*, **25** (2): 387-420.
- Sen, A. (1992). 'Inequality Re-examined'. Oxford: Clarendon Press.
- Siddharthan, K., Ahern, M. and Rosenman, R. (2000). 'Data Envelopment Analysis to determine efficiencies of health maintenance organizations'. *Health Care Management Science* **3**: 23-29.
- Smaje, C. and Le Grand, J. (1997). 'Ethnicity, Equity and the use of health services in the British NHS'. *Social Science & Medicine*, **45** (3): 485-496.
- Smith, P. C. (1992). 'Negative Political Feedback: An examination of the Problem of Modelling Political responses in Public Sector Effectiveness Auditing'. *Accounting Auditing & Accountability Journal*, **5** (1): 5-20.
- Smith, P.C. (1996). 'A Framework for Analysing the Measurement of Outcome'. In P.C. Smith (1996) (eds), 'Measuring Outcome in the Public Sector', London: Taylor and Francis Ltd. 1-19.
- Smith, P.C. and Goddard, M. (2000). 'Reforming health care markets'. In P.C. Smith (eds). 'Reforming markets in health care: an economic perspective'. Buckingham: Open University Press, 1-17.
- Solà, M. and Prior, D. (2001). 'Measuring Productivity and Quality Changes using Data Envelopment Analysis: An Application to Catalan Hospitals'. *Financial Accountability & Management*, **17** (3): 219-245.
- Stewart, J. and Walsh, K. (1994). 'Performance Measurement: When Performance can Never be Finally Defined'. *Public Money and Management*, April-June: 45-49.
- Szczepura, A., Davies, A., Fletcher, C. and Boussofiane, A. (1993). 'Efficiency and Effectiveness in general practice'. *Journal of Management in Medicine*, **7** (5): 36-47.
- Thanassoulis, E., Boussofiane, A. and Dyson, R. G. (1995). 'Exploring output quality targets in the provision of perinatal care in England using data envelopment analysis'. *European Journal of Operational Research* **80**: 588-607.
- The Audit Commission (2000). 'On target: the practice of performance indicators'. London.

- The Department of Health (1997). 'The New NHS: Modern - Dependable', London: The Stationery Office. (Available on http://www.open.gov.uk/The_Department_of_Health/newnhs/newnhs.htm).
- Tsai, P. F. and Molinero, C. M. (2002). 'A variable returns to scale data envelopment analysis model for the joint determination of efficiencies with an example of the UK health service'. *European Journal of Operational Research*, **141**: 21-38.
- Urbanos-Garrido, R. M. (2001). 'Explaining Inequality in the Use of Public Health Care Services: Evidence from Spain'. *Health Care Management Science*, **4**: 143-157.
- Urbina, O., Serrano, D., Marcuello, C. and Garcia, F. (1997). 'Son adecuados los indicadores que se utilizan en Atención Primaria?. La opinión de los profesionales de 3 áreas sanitarias'. *Atención Primaria* **20** (4): 191-94.
- Waters, H. (2000). 'Measuring equity in access to health care'. *Social Science & Medicine*, **51**: 599-621.
- Williams, A. (1974). 'Need' as a demand concept (with special reference to health)', in Culyer, A. (Ed.), 'Economic Policies and Social Goals: Aspects of Public Choice'. London: Martin Robertson.
- Williams, A. (1985). 'The Economics of Coronary Artery Bypass Grafting'. *British Medical Journal*, **291**: 326-329.
- Williams, A. (1996). 'Health and Health Care'. In P.C. Smith (1996) (eds), 'Measuring Outcome in the Public Sector', London: Taylor and Francis Ltd. 20-33.
- Wagner, J. M., Shimshak, D. G. and Novak, M. A. (2003). 'Advances in physician profiling: the use of DEA'. *Socio-Economic Planning Sciences*, **37**: 141-163.
- Zavras, A., Tsakos, G., Economou, C. and Kyriopoulos, J. (2002). 'Using DEA to evaluate Efficiency and Formulate Policy within a Greek National Primary Health Care Network'. *Journal of Medical Studies*, **26** (4): 285-292.