6

Exploring the Use of Data Envelopment Analysis for Evaluation in Primary Care

by

Carla F. Amado

Thesis submitted in partial fulfiment of the requirements for the degree of

Doctor of Philosophy

Supervised by

Prof. Robert G. Dyson

Dr. James E. Storbeck

University of Warwick, Warwick Business School

September 2003

Contents

List of Tables

List of Models

List of Figures

Acknowledgements

There are a number of people that have helped me throughout the four years since I started my PhD. Most credit goes to my supervisors, Professor Robert Dyson and Dr James Storbeck, for their guidance, support and enthusiasm. Special thanks to Robert for his endless words of encouragement and for his prompt comments on improving the thesis, your help was very much appreciated, many thanks. This thesis has also greatly benefited from the comments and discussions during the seminars and other occasions with the Warwick DEA team: Victor Podinovski, Estelle Shale, Joseph Coughlan and Susila Munisamy-Doraisamy.

I would also like to thank the head of the Universidade do Algarve, Professor Adriano Pimpão, Professor Efigénio Rebelo and Professor João Albino Silva, lecturers in this university, for their support throughout this period.

This thesis used data supplied by four Primary Care Trusts in England. Thanks to the information facilitators, Dawn Friend and Sharon Gray, for all your help in the data extraction. I acknowledge financial support from the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia).

I am also very grateful to all my friends whose endless time and encouragement made the concretisation of this work possible. Special thanks to Deborah, Joe, Anna, Amanda, Lucia and Sergio whose enormous help and support throughout these four years cannot be adequately expressed.

I would like to thank my mum, Irene. Obrigada por tudo mãe. And finally, I would like to thank my partner, Lee, for being there for me at all times and providing me with the inspiration to finish this work. Eu amo-te muito.

Declaration

This thesis is my own work and no portion of this work has been submitted in support or application for another degree or qualification of this or any other university or any other institution of learning. This work has not yet been submitted for publication.

The following conference presentations have taken place:

Amado, C. F. (2000), *Performance Measurement in Primary Health Care*, presented at OR42: 42nd annual Conference of the Operational Research Society [in Data Envelopment Analysis and Organisational Performance Measurement], Swansea, Wales, UK, 12/09/00 – 14/09/00.

Amado, C. F. and Storbeck, J. E. (2000), *Policy Interventions and Best Practice*, presented at INFORMS 2000: Annual Conference of the Institute for Operations Research and Management Sciences [in Modelling Regulated Performance], San Antonio, USA, 05/11/00 – 08/11/00.

Amado, C. F. (2001), *Formative Evaluation in Primary Health Care: The Case of Diabetes Care*, presented at: Young OR 12: Conference of the Operational Society [in Data Envelopment Analysis and Organisational Performance Measurement], Nottingham, UK, 27/03/01 – 29/03/01.

Amado, C. F. (2001), *Formative Evaluation in Primary Health Care: Towards a Critical Realist Use of DEA*, presented at 7EWEPA: 7th European Workshop in Efficiency and Productivity Analysis, Conference of the Productivity Analysis Research Network, [in Non-parametric Applications] Oviedo, Spain, 25/09/01 – 27/09/01.

Amado, C. F., Storbeck, J. E. and Ye, H. (2002), *Envelopment and Representation in DEA*, presented at IFORS 2002: 16th Triennial Conference of the International Federation of Operational Research Societies, hosted by the UK Operational Research Society, Edinburgh, Scotland, 08/07/02 – 12/07/02.

Amado, C. F. (2002), *Diabetes Service Delivery - Towards a critical application of DEA*, presented at IFORS 2002: 16th Triennial Conference of the International Federation of Operational Research Societies, hosted by the UK Operational Research Society, Edinburgh, Scotland, 08/07/02 – 12/07/02.

Amado, C. F. (2003), *Using to Measure Efficiency and Effectiveness in Diabetes Care Delivery*, presented at Young OR13: Conference of the Operational Society [in Data Envelopment Analysis and Organisational Performance Measurement], Bath, UK, 2-4 April 2003.

Abstract

Primary care is currently at the heart of the National Health Service policy. The newly established organisations, known as Primary Care Groups and Trusts (PCG/Ts), have the crucial role of improving the efficiency and equity in primary care delivery. However, not many studies have focused on performance assessment in primary care provision.

In this thesis 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. Following a review of the literature, a conceptual framework for performance assessment in primary care was developed in collaboration with a sample of PCG/Ts. This framework 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. DEA is then 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.

Following the Government's policy of focusing on the management of chronic diseases, this conceptual framework is then applied to compare the performance of a sample of GP surgeries in terms of their delivery of diabetes care. An exploratory and formative comparison is undertaken in order to investigate why certain surgeries appear to perform better than others, and in order to identify ways in which some of the surgeries can improve their performance. The relationships between the different performance criteria are also investigated, together with the ways in which an adequate balance between the different criteria can be achieved.

Several conclusions are arrived at and several contributions are made to this research area. The usefulness of complementing efficiency analysis with equity and effectiveness analysis; the usefulness of following a formative methodology in the use of DEA; the importance of estimating the costs involved in achieving speciality based outcomes; the importance of area deprivation in the achievement of effectiveness and the limitations of the data available in primary care in England.

Abbreviations

ACEI	Angiotensin-converting Enzyme Inhibitors
AE	Allocative efficiency
BCC	Banker, Charnes and Cooper
BP	Blood Pressure
CCR	Cooper, Charnes and Rhodes
CCE	Chance Constrained Efficiency
CE	Chief Executive
CHD	Coronary Heart Disease
CRS	Constant Returns to Scale
DCCT	The Diabetes Control and Complications Trial Study Group
DEA	Data Envelopment Analysis
DETR	Department of the Environment, Transport and the Regions
DSQ	Diabetes Service Questionnaire
DRS	Decreasing Returns to Scale
DHA	District Health Authority
DMU	Decision Making Unit
FHSA	Family Health Service Authority
GMS	General Medical Services
GP	General Practitioner
HA	Health Authority
HbA1c	Glycated Haemoglobin
HImP	Health Improvement Plan
HMO	Health Maintenance Organisation
IHD	Ischaemic Heart Disease
IMD	Indices of Multiple Deprivation
IRS	Increasing Returns to Scale
MC	Minimum Cost
NHS	National Health Service
NHSE	National Health Service Executive
NSF	National Service Framework

OE	Overall Efficiency
PACT	Prescribing Analysis and Cost
PAF	Performance assessment Framework
PCC	Primary Care Centre
PCG	Primary Care Group
PCT	Primary Care Trust
PCP	Primary Care Physicians
PHCC	Primary Health Care Centre
PI	Performance Indicator
PMS	Personal Medical Services
PPA	Prescribing Price Authority
PPQ	Practice Profile Questionnaire
PRIMIS	Primary Care Information System
QALYS	Quality-Adjusted Life Years
RHA	Regional Health Authority
SE	Scale Efficiency
SFA	Stochastic Frontier Analysis
TE	Technical Efficiency
UK	United Kingdom
UKPDS	The UK Prospective Diabetes Study Group
USA	United States of America
VRS	Variable Returns to Scale
WR	Weight Restriction
WTE	Whole Time Equivalent

Preamble

"Even if one accepts that there may be some generic management principles that are 'universally' applicable (and that requires a considerable leap of faith) they are always applied in a specific context, including a value context. The nature of the management task, and the appropriateness of the management method deployed, can be defined only in relation to the objectives of the organization being managed, the values to be upheld by its managers as determined by its governors and the status of its relationships with its users, whether citizens, clients, consumers or customers." (Du Gay, 2000: 7)

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 has 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 heath 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 thesis 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. In this respect, we compared a sample of GP surgeries in England in terms of their delivery of diabetes services. In order to fully involve all the GP surgeries, and in order to achieve a high level of confidence in the data and its comparability, a small sample of surgeries was compared. Furthermore, this allowed us to develop more specific performance measures, and to collect the data for its measurement within our limited time period. We used 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.

The broad aims for this thesis are three fold. Firstly, by developing a comprehensive framework to assess the performance of primary care providers, we aim to contribute to the literature on this under-researched area. Secondly, by applying 'Data Envelopment

Analysis' (DEA), a frontier technique for performance measurement, to primary health care in a formative way, we aim to assess its potential to contribute to performance improvement in practice. Past studies using DEA have focused on performance rates, rankings and targets, whereas we use DEA to identify key learning networks, and to design strategies for performance improvement. Thirdly, by undertaking in-depth analyses of the performance results, we aim to draw some conclusions for primary diabetes care delivery, management and policymaking. To be more specific, we have the following research objectives:

1.2 Research Objectives

(1) 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.

(2) To perform comparative performance analyses at the level of GP surgeries in England, in order to provide an assessment of the complexities involved in measuring and managing performance in primary care. The specific objectives of this empirical case study are:

- To apply the conceptual generic framework developed to diabetes care delivery;
- To analyse the revealed implicit trade-offs between the different dimensions of performance;
- To analyse the impact of non-controllable factors, such as patient compliance and area deprivation on the performance results;
- To explore the usefulness of the DEA results in practice, for example, in order to establish learning networks between the surgeries;

- To analyse the pattern of results to explore some of the potential causes behind the differences in performance. This should identify some of the factors behind successful performance in primary diabetes care delivery.

1.3 Context – The National Health Service and Primary Care Delivery in the UK

Since its creation in 1948, the NHS is one of the most important public organisations in the UK. It was created to provide every citizen with comprehensive high quality health care, free at the point of consumption. This single organisation was created to satisfy the health care needs of the population, which were expected to decrease over time, assuming that the population would become healthier (Ham 1999). The fact that the health care needs have not stopped increasing has caused enormous financial pressure on this organisation and has motivated successive reforms of the NHS.

As pointed out by Levitt and Wall (1984: 10), the final form of the NHS established in 1948 resulted from a compromise between conflicting demands of the several interest groups, and therefore it was not a *unified health service* providing free comprehensive health services to every citizen, as initially intended. Rather, a system characterised by the fragmented delivery of hospital, community and Family Practitioners care resulted.

1.3.1 Attempts to Unify the NHS

After 1948, several efforts were made to unify the NHS, which are visible in the published reports and the parliamentary acts passed (Levitt and Wall 1984). However, the different interest groups could not agree on a unified NHS. In particular, GPs concerned with their secondary role in the NHS, felt that unification would not satisfy their interests. It was not until April 1, 1974 that a new structure for the NHS was put into practice. This reorganisation's main objective was to improve the cooperation between the hospital, community and Family Practitioners. It was hoped that by working cooperatively, the efficiency and quality of care in the NHS would also improve. In this respect, a single administrative unit managing all three parts of the NHS needed to be established. Regional Health Authorities (RHAs) were created to unify the strategic planning of the NHS at a regional level, whilst newly established Area Health Authorities (AHAs) had the task of planning and managing the services provision at local level.

This new structure led to an increase in the administration costs, and such an expensive structure characterised by multiple tiers of administration was soon to be the cause of criticisms (Levitt and Wall 1984). Stronger delegation of powers for decision-making was seen as desirable. On April 1, 1982, all 90 AHAs were removed and 192 District Health Authorities (DHAs) were introduced.

1.3.2 The Influence of New Public Management

During the late 1980s, a more radical set of reforms was undertaken by the Conservative Government of Margaret Thatcher. These reforms culminated with the publication of the White Paper, *Working for Patients* (Secretary of State of Health and Others 1989a). The Government intended to bring the principles of management within the private sector into the administration of the public sector. This style of public administration became known as the *New Public Management* (Hood 1995; Du Gay 2000) and is characterised by certain elements: (1) changing the identity of public organisations by introducing entrepreneurial identity and practice. This leads to management by leadership, with clear missions and visions, as opposed to rules and regulations; (2) managerial decentralisation, encouraging participative management with an associated increase in upwards accountability in terms of achieved results; (3) the use of market-type mechanisms and practices, and the introduction of competition between service providers; (4) redefinition of citizens as customers, and orienting service provision towards increasing customer satisfaction; (5) the use of performance assessment as a way of improving services, with a strong focus on outcomes achievement; (6) using partnerships between the public, private and voluntary sector in order to improve results and (7) empowering citizens and communities to take control and responsibility in public services.

Most of these ideas were already being applied to the NHS at the beginning of the 1980s. The publication of a list of Performance Indicators (PIs) in 1983 can be considered the first visible sign of these ideological reforms. However, the introduction of the internal market in 1991 represents the most significant modification on the operation of the NHS. The internal market was characterised by the separation between service provision and purchasing. On the provider side, self-governing NHS Trusts were created to manage and deliver hospital services. On the purchaser side, DHAs and Family Health Services Authorities, together with some General Practices (known as fundholders), were assigned a budget to purchase health services. This relationship between purchasers and providers was formally stipulated in contracts, and aimed at improving the efficiency of service delivery. The introduction of fundholding (devolved budgets to General Practices) was intended to improve the performance and accountability of the doctors, by involving GPs in the management of the assigned budget.

The fundholding scheme proved to be quite appealing, by 1997 half of the population was registered with a fundholding practice (Smith and Goddard 2000). However, the fact that fundholding was not introduced on a compulsory basis created doubts about equity in the service provision. It was suggested that patients belonging to a fundholding practice were benefiting from better access to services when compared to patients from a non-fundholding practice (Le Grand, Mays and Mulligan 1998).

The main aim of the Conservative Government reforms of 1989 was to increase the efficiency of the hospital services, always assumed to be the most important part of the NHS. The introduction of fundholding in 1990, on a voluntary basis, was more a consequence of this initial aim than an aim in itself (Bloor, Maynard and Street 2000). Furthermore, the accountability structure regarding the delivery of Family Practitioners services has also been unclear, and its independent contractor status gives GPs an autonomous role. It is only in 1997 that the role of primary care is explicitly considered in another set of health system reforms introduced by the Labour Government.

1.3.3 After 1997: A Primary Care-led NHS?

The White Paper, *The New NHS: Modern and Dependable* (1997), defines the objectives of yet another reform on the NHS. According to this document, this is meant to be "a 'third way' of running the NHS - a system based on partnership and driven by performance" (The Department of Health 1997, Paragraph 2.2). Concerns regarding inequity of service provision derived from the fundholding system gave rise to significant alterations of the organisation of primary health care. For the first time, primary care has been considered as central to the success of the reforms of the NHS (The Department of Health 1997). This shift of emphasis from hospital services to primary care is one of the major differences of this reform, along with a theoretical emphasis on the equity of service provision.

A definition of what constitutes primary health care is essential before we proceed. The National Health Service Executive (NHSE) defines it as: "first contact, continuous, comprehensive, co-ordinated care provided to individuals and populations undifferentiated by age, sex, disease or organ system" (NHSE 1997, as quoted in Dowell and Neal 2000: 10). Primary care is the first line of support for patients, as most patients are registered with one GP surgery. Furthermore, with the exception of emergencies; patients gain access to the NHS through primary care. In terms of resources used, primary care controls a significant part of the NHS budget both directly, through the delivery of primary care services and indirectly, through referrals to community care and secondary care. General practice is part of primary care. It involves the services delivered by GPs, practice nurses and other staff who work in GP surgeries.

The 1997 reforms instituted by the Labour Government created a new figure in primary health care: Primary Care Groups (PCGs). 481 PCGs were established in April 1999 across England to commission health services for patients. PCGs are composed of a set of individual health practices grouped together on a regional basis, covering a population of around 100,000 patients. Each PCG is headed by a Board (governing body) constituted by four to seven GPs; one or two community or practice nurses; one Social Services nominee; one lay member; one Health Authority (HA) non-executive member and one PCG Chief Executive. They have access to a unified budget to commission and provide health services, and are accountable to justify how they utilize resources.

The early statements from the Government about PCGs' Board composition portrayed a pluralistic view of decision-making, with an emphasis on fair representation from all the relevant stakeholders in primary care. However, following a long established tradition, the GPs threatened not to cooperate unless certain demands were satisfied, which implied a shift towards a GP dominant Board (Smith and Sheaff 2000). With the current structure, GPs have a clear majority in the Boards, with the legal assurance that no meeting can start without a GP majority. This is an important aspect as it assures that the decision-making process in the PCGs is dominated by the GPs' values and priorities. Moreover, GPs have preferential status in terms of the way they are elected to the Board: while GPs are peer elected, the other members have to be selected by the HA.

PCGs were designed to develop over four stages. In stages one and two they only have an advisory role towards the HA and, in stages three and four, they become established as freestanding bodies accountable to the HA. At the last two stages they are called Primary Care Trusts (PCTs) and have the responsibility of fully managing unified cash limited budgets to cover General Medical Services (GMS) infrastructure, prescribing, hospital and community services. The HA decides when a PCG (level 2) is ready to upgrade to PCT, upon a formal application from the PCG. The objective is that every PCG will upgrade to PCT status before 2005. PCTs are governed by a chairman and lay members appointed by the Secretary of State, and managed by a Chief Executive, a finance director and three other executive members.

According to the White paper *The New NHS* (The Department of Health 1997: Paragraph 5.9.), the main functions of these new primary care organisations are to:

(1) contribute to the development of the annual Health Improvement Plan (HImP);

(2) promote the health of the local population;

(3) commission health services for the local population;

(4) monitor the performance of NHS trusts delivering services for the local population;

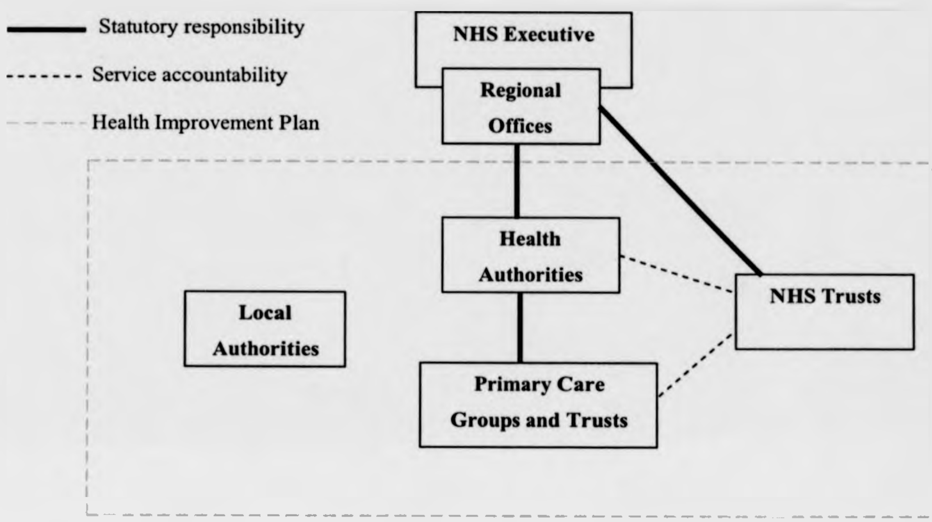
(5) develop primary care;

(6) better integrate primary and community health services.

The transference of the operational role in primary care delivery from HAs to PCG/Ts is an essential feature of the Labour Government reforms. A more strategic role was given to the HAs. In collaboration with Local Authorities, PCG/Ts and NHS Trusts, the HAs were given the task of developing HImPs to identify the needs of the population and the most adequate strategies to satisfy these needs. HAs are also expected to set management and service partnerships with other organisations inside and outside the NHS. The role of performance monitoring is also an important part of HAs tasks. They allocate resources to its PCG/Ts and set up local performance targets, holding PCG/Ts accountable to the achievement of these targets. PCG/Ts set indicative budgets for their individual practices, and have to develop mechanisms to ensure that budgets are not exceeded (Bloor, Maynard and Street 2000). Figure 1.1 represents the current organisational structure of the NHS, with the several lines of accountability in operation.

Figure 1.1: NHS accountability structure from 1997 (Source: The Department of Health 1997)

Key:



In practice, it can be argued that these latest reforms show strong signs of a continuation of the previous policies and models, with a particular emphasis on applying the principles of *New Public Management* in primary care. This is indicated by the introduction of the concept of clinical governance in primary care.

> "Clinical governance can be defined as a framework through which NHS organisations are accountable for continuously improving the quality of their services and safeguarding high standards of care by creating an environment in which excellence in clinical care will flourish."
>
> (The Department of Health 1997: Paragraph 3.2.)

According to the concept of clinical governance, performance assessment is a vehicle for quality improvement and it is the responsibility of each organisation and each professional in the NHS to work towards the continuous improvement of quality. This concept of clinical governance has clear relationships with the concept of 'entrepreneurial governance' in the private sector.

> "Performance management and related techniques, for example, involve a characteristically 'contractual' relationship between individual employees and the organization for which they work. This involves offering individuals involvement in activities – such as managing budgets, training staff, delivering services – previously held to be the responsibility of other agents – such as supervisors, personnel departments and so forth. The price for this involvement is that individuals themselves must assume responsibility for carrying out those activities and for their outcomes." (Du Gay 2000: 85)

To put clinical governance into practice, a Performance Assessment Framework (PAF) was developed (National Health Service Executive 1999). PAF measures performance upon six dimensions: (1) health improvement, (2) fair access, (3) effective delivery of appropriate health care, (4) efficiency, (5) patient and carer experience of the NHS and (6) health outcomes of NHS care. From this framework, a list of PIs was published to assess the performance of HAs. The list of PIs defined in PAF is presented in Appendix A. However, how HAs and PCG/Ts are supposed to learn from these indicators in order to improve the quality of services is not clear. It is therefore crucial to undertake a formative evaluation in primary care aimed at performance improvement. We will next justify why we believe that Data Envelopment Analysis is an appropriate technique to compare primary care providers.

1.4 Why use DEA for Performance Assessment in Primary Health Care?

One method to assess performance is to benchmark the organisation against others performing similar activities. A benchmarking exercise allows the organisation to evaluate how well it is doing, when compared to best-observed practice according to a particular criterion. Also, a good benchmarking exercise should allow the organisation to learn about *how* other organisations are able to do better. However, the challenge is in identifying 'best practice', as what constitutes 'best practice' is in itself dependent upon the perspective taken. Furthermore, what constitutes 'best practice' is dependent upon the priorities of the organisation under evaluation. Therefore, what constitutes 'best practice' to one organisation might not constitute 'best practice' to another. If one health care delivery unit, responding to its local needs, gives priority to the delivery of preventive services, its performance should preferentially be compared with the performance of units with similar priorities.

DEA is a non-parametric technique used to measure the efficiency of Decision Making Units (DMUs) and was first proposed by Charnes, Cooper and Rhodes (1978). This technique uses a production metaphor. It considers that each DMU is engaged in a transformation process, where by using some inputs (resources) it is trying to produce some outputs (goods or services). One of the interesting features of DEA is that it allows each unit to identify a benchmarking group; that is, a group of units that are following the same objectives and priorities, but performing better. In this respect DEA aims to respect the priorities of each DMU by allowing each one of them to choose the weight structure for inputs and outputs that most benefits its evaluation. As a result, it aims to classify each unit in the best possible light.

There are several distinct reasons that make DEA suitable for use in performance assessment in primary health care:

(1) DEA can handle the existence of several non-commensurate inputs and outputs, without requiring unit prices for each of them. This is a clear advantage in health care studies, when we consider that we are usually confronted with processes characterised by the use of several resources, involved in the delivery of several distinct services, in order to achieve several distinct outcomes. Whilst unit costs for the resources may be obtainable, we do not have information regarding the unit value of outputs and outcomes;

(2) DEA uses all the data available to construct a best practice empirical frontier, to which each non-optimal production point is compared. This is a distinct advantage when compared to regression-based techniques that focus on the average units and construct an average production function;

(3) In contrast with parametric techniques, such as Stochastic Frontier Analysis (SFA), DEA does not require the specification of the functional form that links the inputs to the outputs. Moreover, if information is available, it allows the incorporation of several distinct assumptions regarding the relationships operating at the frontier, regarding the returns to scale, the disposability of the inputs and outputs, and the convexity of the production function;

(4) The performance of DMUs can be assessed by using several alternative orientations to the best practice frontier, depending on the context of the study. The DEA projections to the best practice frontier can reflect an output maximisation, an input minimisation, or a combination of both;

(5) Given that most DEA models can be conceptualised as a linear program, a vast range of useful information becomes available for research, decision-making and policy making.

Despite these important advantages, we are fully aware of the limitations of this technique. These limitations include the following:

(1) DEA assumes that it is possible to fully characterise the production of health care by identifying a set of inputs, outputs and outcomes of production. However, some of the outputs and outcomes of primary care are not measurable. Furthermore, there are an extensive number of factors operating at different levels which impact on this transformation and which may lead to the violation of the linear input-output relationship. Awareness of these factors and of their potential impact on the results is necessary if useful information is to be obtained;

(2) DEA assumes one-way causality between inputs to outputs. In this respect, it is assumed that higher levels of inputs should lead to higher levels of outputs. However as pointed out by Smith (1992), in a context where the level of output achievement is published, poor levels of outputs can generate pressure for increased resources. In this respect, a negative correlation may exist between inputs and outputs. Smith (1992) refers to this problem as 'negative political feedback'. Orme and Smith (1996) examined the effects of this problem for DEA studies and concluded that, in the presence of 'negative political feedback', DEA estimates of efficiency may be subject to bias against DMUs using fewer resources. These DMUs may be asked to make greater improvements in efficiency than DMUs using higher levels of resources. They further concluded that this problem is more pronounced in small samples. In this respect, the use of DEA for target setting in a context where the output data is subject to great political scrutiny requires further caution.

(3) Measuring outputs represents a utilitarian approach that values the consequences of actions above anything else. In this respect, the use of DEA to assess performance leads to a *biased* representation of success in organisations. The delivery of public services is also governed by non-consequentalist principles and it is essential to provide alternative representations of success.

(4) Given the arbitrariness involved in the selection of variables used in a DEA model (particularly output and outcome variables), a DEA analysis can lead to biased and potentially misleading results. For example, there is a danger that managers could suggest the introduction of variables that they know will benefit their classification. This is one of the reasons why we suggest the use of DEA for formative evaluation;

(5) DEA does not take into account stochastic variability in the data and is therefore very sensitive to data errors;

(6) DEA assumes proper envelopment of all the DMUs. That is, a meaningful measure of efficiency can only be calculated if the DMU under evaluation has a comparison set. Otherwise, the DMU will be classified as efficient, simply because no comparison can be established. This is a limitation that results from the fact that DEA allows full flexibility in the weights applied to each input and output. In this respect, if a DMU is characterised by a peculiar mix of inputs or outputs, it can find a set of very extreme weights under which it is efficient. This problem can be solved by the introduction of weight restrictions regarding the admissible ranges of input or output weights.

1.5 Thesis Outline

This introductory chapter provided an overview of the thesis and of its objectives. A brief introduction to the context of study was included, with particular emphasis to the reforms that have been implemented in the NHS since its creation. The latest reforms of the NHS were discussed in relation to the importance given to primary care in the delivery and coordination of health services. The introduction of PCG/Ts, their functions, and the new strategic role given to HAs were discussed. The importance of performance assessment as part of the Government's modernisation agenda was discussed together with the suitability of DEA for performance assessment in primary care.

Chapter 2 provides an introduction to 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. The need to include weight restrictions in order to obtain meaningful results is pointed out and several approaches for their inclusion in DEA are discussed.

Chapter 3 reviews the relevant literature. We review the methods that can be used to compare primary health care providers and the empirical studies that have been carried out. By reviewing the performance assessment studies in primary health care, we identify the main issues that still need to be addressed.

Chapter 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 (diabetes care) is justified. The advantages and limitations of comparing primary care providers based on a speciality are also discussed.

Chapter 5 discusses the methodology used to develop the models and, to collect and analyse the data. In this chapter we justify the importance of using a combination of research methods in order to collect, analyse and interpret the performance data. The importance of developing a clear conceptual framework regarding the context of study is emphasised as a way of giving coherence to the project. We also discuss the development of two questionnaires to collect data that was not available. This is followed by a discussion of the design of each one of the questionnaires and the methodology used for sample size calculation.

Chapter 6 provides the background for the empirical study. We start by justifying the need for a formative evaluation in primary diabetes care and briefly characterising this chronic condition. We then return to the conceptual performance assessment framework as presented in chapter 4 and discuss its applicability in primary diabetes care delivery. In particular, we discuss the local needs, objectives, resources used, services delivered and outcomes achieved in primary diabetes care. These are the key elements for the development of the performance assessment models presented in chapter 7.

Chapter 7 presents the data, the models and the results of the primary diabetes care evaluation undertaken. We discuss the data available and its quality. A brief characterisation of the surgeries under evaluation is performed in order to provide a background for the results. The results in terms of a comparative analysis of equity, efficiency, clinical effectiveness, patient-focused effectiveness and cost effectiveness are presented and discussed.

In chapter 8 we analyse the trade-offs between the different performance criteria and conclude that the performance status of some surgeries varies greatly with the criterion chosen. We conclude chapter 8 by assessing the robustness of the performance results to the introduction of additional surgeries.

Chapter 9 discusses the effects of the environmental variables on the performance results. The visible effects of area related factors and patient related factors are discussed. We also discuss the effect of different structures and processes of care in the performance results. We report on the workshops undertaken with the PCTs and the surgeries to discuss the use of the results in practice.

Chapter 10 provides a summary of the thesis and the main conclusions of the work. It discusses the main contributions of this study for research, decision-making and policy making. A discussion of the limitations of this study and suggestions for further research is also included.

2 An introduction to Data Envelopment Analysis

2.1 Introduction

In this chapter we introduce the technique used 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. We introduce the concept of weight restrictions and discuss their interpretation in section 2.3.

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.

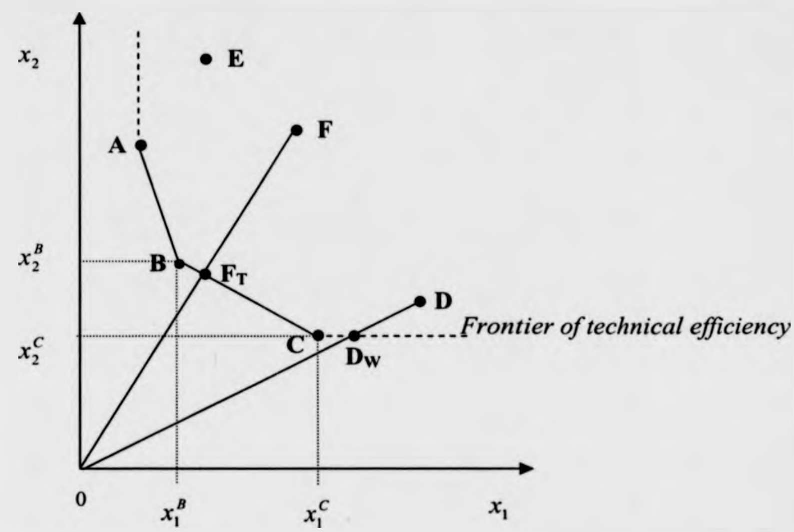
2.2.1 Graphical Representation of DEA – Radial and Non-radial Efficiency Measures

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:

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

Figure 2.1: The measurement of technical efficiency



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:

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

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.

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, $j = 1, \ldots, n$) to be compared, which use m inputs (x_i, $i = 1, \ldots, m$) to produce s outputs (y_r, $r = 1, \ldots s$). 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:

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

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

$$v = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \qquad u = \begin{bmatrix} u_1 & u_2 & \cdots & u_s \end{bmatrix}$$

We can calculate the technical efficiency (*TE*) of DMU_0 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
$TE_0(x,y)^{CRS} = Min\ \theta$	$TE_0(x,y)^{CRS} = Max\ uy_0$
Subject to:	*Subject to:*
$\theta x_0 - X\lambda \geq 0$	$vx_0 = 1$
$Y\lambda \geq y_0$	$-vX + uY \leq 0$
$\lambda \geq 0.$	$v \geq 0,\ u \geq 0.$

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':

$$s^- = \theta x_0 - X\lambda,\ s^+ = Y\lambda - y_0$$

Also, let us define a vector of ones:

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

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:

$$\theta^* = 1 \text{ and } s_i^{-*} = 0 \; \textit{(for all i, i = 1, ..., m)}$$

$$\text{and } s_r^{+*} = 0 \; \textit{(for all r, r = 1, ..., s)}$$

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

$$Max \;\; \omega = es^- + es^+$$

Subject to:

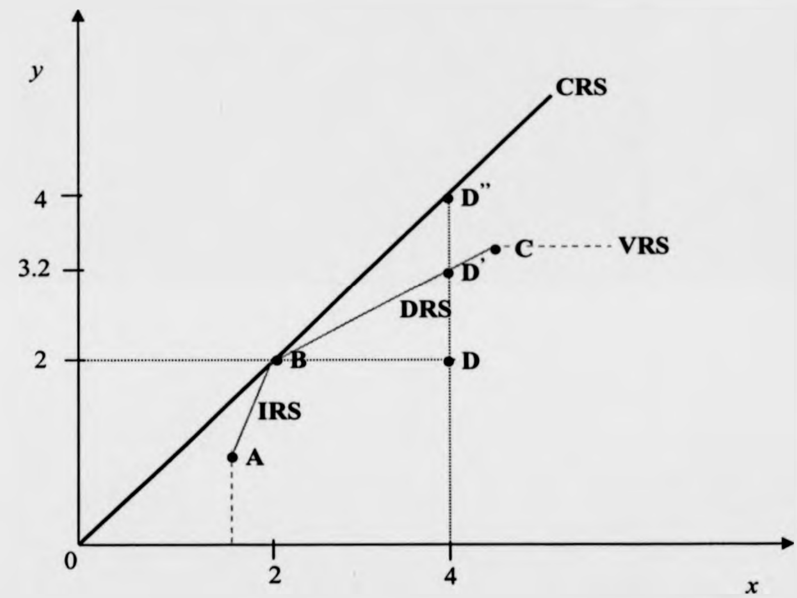
$$s^- = \theta^* x_0 - X\lambda$$

$$s^+ = Y\lambda - y_0$$

$$\lambda \geq 0, \;\; s^- \geq 0, \;\; s^+ \geq 0.$$

Banker, Charnes and Cooper (1984) proposed a DEA model that assumes the existence of variable returns to scale (VRS) and is applicable when returns to scale exist. In order to introduce the concept of returns to scale and its implications for efficiency measurement, let us introduce another simple example. Figure 2.2 presents 4 DMUs (A, B, C and D), which use one input (x) to produce one output (y).

Figure 2.2: Productions frontiers under CRS and VRS (unique most productive scale size)



In this simple example, two frontiers of efficiency have been constructed. One frontier assumes CRS and is represented by the line in bold that passes through the origin. The other frontier assumes the existence of VRS and is the line that links point A with B and point B with C. The vertical extension from point A and the horizontal extension from point C constitute the dominated parts of the VRS frontier.

DMU B is the *only* unit deemed efficient under the assumption of CRS, because it is the only unit located in the CRS frontier. Therefore, DMU B is the only unit operating under the most productive scale size. All the other units operate under sub optimal scale sizes. DMU A operates at increasing returns to scale, i.e., a greater scale of operation would lead to a greater ratio of outputs over inputs; DMU C operates under decreasing returns to scale, which means that a smaller scale of operation would lead to a greater ratio of outputs over inputs. DMU D is inefficient under any assumption regarding returns to scale. However, as shown graphically, its efficiency measure varies with different returns to scale assumptions. If an output orientation is used, the efficiency of DMU D under CRS and under VRS is respectively:

$$TE_D(x, y)^{CRS} = \frac{OD}{OD''} = 0.5$$

$$TE_D(x, y)^{VRS} = \frac{OD}{OD'} = 0.625$$

As it can be easily understood, the efficiency measure under CRS is smaller than the efficiency measure under VRS. In fact, in any situation, we have the following relationship:

$$TE(x, y)^{CRS} \leq TE(x, y)^{VRS}.$$

It is also important to note that the efficiency measure under CRS remains the same independently of the orientation taken. However, under VRS this relationship does not necessarily hold. In the example shown, if an input orientation is used, the efficiency measure for DMU D under VRS is the same as the efficiency measure under CRS, which is 0.5. However, as previously stated, with an output-orientation, the efficiency measure under VRS is 0.625.

The assumption of VRS imposes an additional restriction in the envelopment form of the DEA model. In particular, we impose that the DMU under evaluation can only be projected to a convex combination of DMUs *of the same scale*. I.e., we require that the sum of the intensity variables be equal to one ($\Sigma_j \lambda_j = 1$). Banker, Charnes and Cooper (1984) developed a DEA model to measure technical efficiency assuming VRS. This model is known as the BCC model and is presented in Model 2.3.

Model 2.3: The BCC Model – input oriented

$$TE_0\,(x, y)^{VRS} = Min\ \theta$$

Subject to:

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

$$Y\lambda \geq y_0$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda \geq 0.$$

Banker, Charnes and Cooper (1984) showed that the CCR measure of technical efficiency (assuming CRS) can be decomposed into the multiplication of a pure technical efficiency measure obtained from the BCC model (assuming VRS) and a scale efficiency measure. I.e.:

$$TE_0\,(x, y)^{CRS} = TE_0\,(x, y)^{VRS} \times SE_0\,(x, y)$$

In this respect, if we want to calculate a measure of scale efficiency for DMU_0 (SE_0), we use the following formula:

$$SE_0\,(x, y) = TE_0\,(x, y)^{CRS} / TE_0\,(x, y)^{VRS}, \text{ with } SE_0\,(x, y) \leq 1.$$

The scale efficiency of a DMU is equal to 1 when the DMU is CRS efficient. If the DMU is VRS efficient, but not CRS efficient, then its scale efficiency is smaller than one, indicating that the DMU is not operating at the most productive scale size.

The choice over the type of scale assumption used should be contextual and should result from an analysis regarding the type of inputs used and the type of outputs used. If there is evidence that changes in the inputs should not lead to proportional changes in the outputs, then a VRS assumption should be used; otherwise, the CRS assumption should be used. However, it is equally important that the VRS model is not used when there is no evidence of scale effects.

2.2.2.1 Input-oriented versus Output-oriented Models

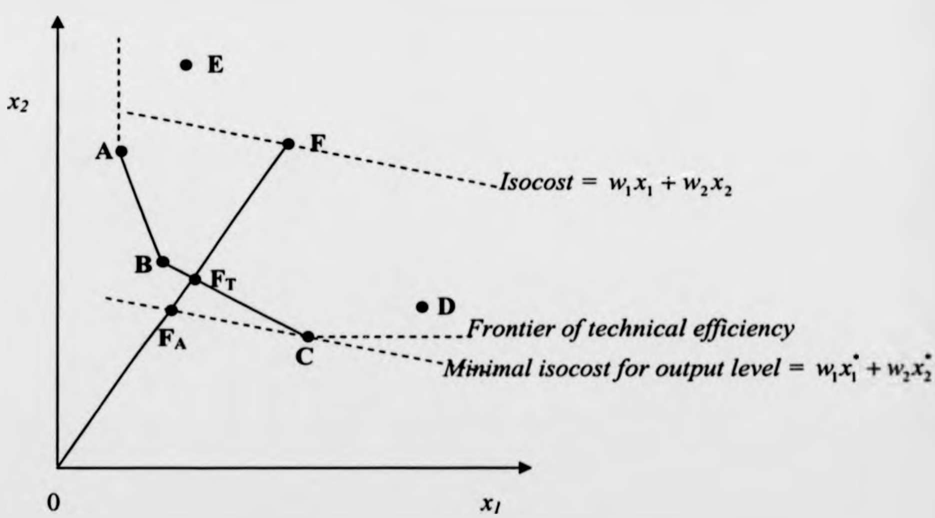
The term 'input-orientation' defines the objective of the envelopment problem, which aims to minimise the value of inputs used in the transformation process. I.e., it looks for the value θ^*, which ensures the maximal radial reduction in the input vector. In a similar way, the term 'output-orientation' defines the objective of maximising the radial augmentation in the output vector.

The choice between using an input-oriented model or an output-oriented model must first of all reflect the control the decision-maker has over each one of these objectives. For example, if we consider as an output "the number of patients registered with a GP", an output-oriented model would not provide meaningful results because the decision-maker cannot have full control over this output. If the outputs and inputs are such that in the short-term the decision-maker has full control over both types of variables, then the orientation should reflect the current priorities of the units under evaluation, or an additive model (Cooper, Seiford and Tone 2000) could be used if both objectives are sought.

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 2.3 represents 6 DMUs (A, B, C, D, E and F) using two inputs (x_1 and x_2) to produced 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).

Figure 2.3: The measurement of technical, allocative and cost efficiency



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 allocativelly 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 allocativelly efficient because it uses the optimal input mix given the current inputs 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 allocativelly inefficient. Its measure of allocative efficiency is calculated as:

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

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

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

A DEA model can be formulated to measure allocative efficiency. Let us define a price vector containing the unit prices of each input: $w = [w_1\ w_2 \ldots w_m]$. Assuming constant returns to scale and taking an input orientation, the minimum cost (*MC*) of production for $\mathrm{DMU_0}$ can be calculated by solving the Linear Program presented in Model 2.4 (Färe, Grosskopf and Lovell 1994).

Model 2.4: The Cost Minimisation Model

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

Subject to:

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

$$Y\lambda \geq y_0$$

$$\lambda \geq 0.$$

Model 2.4 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 2.4 the quantities of inputs are variables to be optimised. Once the optimum quantities of inputs have been identified, the minimum feasible cost for DMU_0 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 DMU_0 can be produced, considering the current input prices. If DMU_0 is cost efficient, then its current cost $(w_0 x_0)$ 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

$$\dot{x_0} = \left[\dot{x_{10}} \quad \dot{x_{20}} \quad \cdots \quad \dot{x_{m0}}\right]^T$$

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):

$$\textit{Overall or cost efficiency} = OE_0\,(x, y, w) = \frac{MC_0(x,y,w)}{w_0 x_0}.$$

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 2.4, and calculate the allocative component using the following formula (Färe, Grosskopf and Lovell 1994):

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

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

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

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 discuss the different approaches that have been proposed to introduce weight restrictions in DEA in order to justify the approach used in our case study.

2.3 Introducing Weight Restrictions in DEA

In the previous sections we have discussed the use of DEA to calculate measures of relative efficiency in the presence of multiple inputs and multiple outputs. One of the great advantages of DEA is that it produces efficiency measures without the need to specify the weights associated with each input and with each output. DMU specific weights are calculated during the optimisation process in order to produce the highest possible efficiency measure. However, if total flexibility of the weights is allowed, it is very common to find results where most of the variables have been given a zero weight. This situation is unsatisfactory, given that if the optimal weight associated with a certain input is zero, this means that this input was excluded from the efficiency measurement.

The ratio between the weights assigned to two inputs or two outputs should reflect realistic production trade-offs. In that respect we argue that, whenever possible, information regarding realistic production trade-offs should be included in the DEA model in order to improve the economic validity of the results. For example, if we know that one senior doctor can be substituted by two or more training doctors, without any effect on the services delivered, then this information should be included in the DEA model. This information is introduced in a DEA model via the introduction of restrictions on the weights assigned to the different inputs and outputs (please refer to the multiplier form of the CCR model presented in Model 2.1).

Several approaches have been suggested regarding the best way to formulate and introduce weight restrictions (WR) in a DEA model. We will now review these approaches in order to justify the one that we have followed in this thesis.

2.3.1 The Assurance Region Approach

Thompson et al. (1986) developed the first approach known as the 'assurance region' approach. This approach imposes restrictions regarding the relative magnitude of the weights for certain inputs or outputs. In this respect, it limits the variability of weights between certain inputs or certain outputs. Assurance region restrictions are usually introduced on the multiplier form of the DEA model. For example, if a restriction on the relationship between the weights of inputs 1 and 2 were to be imposed, it would take the following form:

$$L_{1,2} \leq \frac{v_2}{v_1} \leq U_{1,2}$$

2.3.2 Absolute Bounds

Dyson and Thanassoulis (1988) suggested the inclusion of restrictions on the individual inputs and outputs weights, based on economic information regarding the average level of input needed to produce particular outputs. By imposing absolute bounds on the weights, the exclusion of certain inputs or outputs from the efficiency measurement, by assigning them a zero weight, is not allowed. Absolute weight restrictions are usually introduced on the multiplier model and are of the following form:

$$v_i \geq L_i \quad \text{or} \quad v_i \leq U_i \text{ (for inputs)} \quad \text{and} \quad u_r \geq L_r \quad \text{or} \quad u_r \leq U_r \text{ (for outputs)}$$

2.3.3 The Cone-ratio Envelopment Approach

A more general approach to the introduction of weight restrictions has been proposed by Charnes et al. (1990), and is known as the 'cone-ratio envelopment' approach. This approach involves the definition of the admissible area for the weights by using convex cones, spanned by several direction vectors.

The multiplier form of a DEA model after the introduction of restrictions using this approach is readily interpretable: it solves the CCR problem, but it limits the input weights to belong to the cone space defined by the direction vectors determined by the weight restrictions introduced. The envelopment form of the problem can be defined as a problem equivalent to the CCR model on its envelopment form, but with transformed data[1]. This is an important advantage given that once the data has been transformed we can use the standard DEA software available to solve a problem without weight restrictions but with transformed data.

2.3.4 Imposing Restrictions on the Virtuals

Wong and Beasley (1990) proposed a method based on the relative contribution of each input (or output) to the total costs (or benefits) of the DMU. The relative contribution of one input to the global cost is called the virtual weight of that input. Wong and Beasley (1990) suggested the imposition of lower and upper bounds on the virtuals, by means of introducing restrictions of the following type:

---

[1] Cooper, Seiford and Tone (2000: 160, 161) show how the envelopment form is derived from the multiplier form.

$$L_i \leq \frac{v_i x_{ij}}{\sum_{i=1}^{m} v_i x_{ij}} \leq U_i .$$

Pedraja-Chaparro, Salinas-Jimenez and Smith (1997) built on this idea of imposing weights on the virtuals, but suggested that the important aspect is to place restrictions between the virtuals of different inputs (or different outputs). The idea is that the proportion of the costs assigned to a particular input should not be allowed to exceed the proportion of costs assigned to another input by a certain multiple. This multiple is dependent upon the set of inputs and outputs included in the model and should be chosen by the analyst or decision maker. This approach leads to the imposition of weight restrictions of the following type:

$$c_i v_i x_{ij} \leq v_k x_{kj} \leq d_i v_i x_{ij} \text{ with } i \neq k .$$

Some variations of these approaches have also been proposed. Allen et al. (1997) review the literature on the introduction of weight restrictions. Independently of the approach used, the introduction of weight restrictions has been mainly motivated by the need to improve discrimination in the DEA results. However, as pointed out by Podinovski (2002), this has led to a great deal of confusion regarding the meaning of the weight restrictions imposed and the meaning of the efficiency measures obtained from a DEA model with weight restrictions. Some technical problems regarding the validity of the DEA results after the introduction of certain types of weight restrictions have also been highlighted (Podinovski and Athanassopoulos 1998; Podinovski 1999 and Podinovski 2001).

Podinovski (2002) suggests that the confusion that exists regarding the effects of introducing weight restrictions in a DEA model results from the fact that in most approaches the weight restrictions are included in the multiplier form, without an attempt to interpret its effect in the envelopment form. Podinovski (2002) argues that the weight restrictions should be added to the envelopment form of the problem and a clear economic interpretation of their meaning should be sought. He demonstrates that each linear weight restriction, in the multiplier form, generates additional terms in the envelopment form (Podinovski 2002: 3) and then shows that "these terms can be thought of as the trade-offs between the inputs and/or the outputs" (Podinovski 2002:3). Podinovski names his new approach to the introduction of weight restrictions as 'the trade-offs approach'.

The trade-offs approach has several advantages. Firstly, if the weight restrictions are defined based on technologically realistic trade-offs between the variables, the efficiency results will be more meaningful. Secondly, it can be shown that by adding weight restrictions based on trade-offs, *the meaning* of the efficiency results *is preserved* (Podinovski 2002). That is, in an input-oriented model, θ^* means the maximum radial reduction that can be applied to all inputs without worsening the level of outputs produced, exactly the same meaning as for a problem solved without weight restrictions. Thirdly, prior to the introduction of weight restrictions, we can be sure of their feasibility. We shall use the trade-offs approach to introduce weight restrictions in our case study. We will now briefly review this approach.

2.3.5 The Trade-offs Approach

We shall focus only on the inclusion of homogeneous weight restrictions, given that these are the only types of weight restrictions we shall be including in our empirical case study. We will use the notation introduced by Podinovski (2002) and present an example regarding the introduction of a homogeneous weight restriction. Let us take a general homogeneous weight restriction of the form:

$$a_1u_1 + a_2u_2 + \ldots + a_su_s - b_1v_1 - b_2v_2 - \ldots - b_mv_m \leq 0. \qquad (1)$$

Where each of the coefficients a_r *and* b_i can be positive, zero or negative. For example, let us consider the following pair of weight restrictions:

$$1 \leq \frac{v_2}{v_1} \leq 2 \qquad (2)$$

These two restrictions can be translated into two linear equations that fit the general form presented in (1):

$$v_1 - v_2 \leq 0$$

$$-2v_1 + v_2 \leq 0.$$

Podinovski (2002) shows that if each one of these linear equations is a restriction in the multiplier form of the DEA model, it gives rise to a new term ρ in the envelopment form. In particular, for each weight restriction of the form (1), we can see the following effects on the envelopment form of the DEA model:

(1) A new term $a_r\rho$ is added to each output restriction;

(2) A new term $b_i\rho$ is added to each input restriction.

Model 2.5 below represents the envelopment form of the CCR model, after the inclusion of one generic weight restriction of the form stated in (1).

Model 2.5: The CCR Model after the inclusion of one WR (envelopment form)

$$TE_0\,(x, y)^{\text{CRS, subject to WR}} = Min\ \ \theta$$

$$Subject\ \ to:$$

$$\textstyle\sum_{j=1}^{n} \lambda_j y_{jr} + a_r \rho \geq y_{0r} \qquad \text{(for all } r\text{, with } r = 1,\ldots s\text{)}$$

$$\textstyle\sum_{j=1}^{n} \lambda_j x_{ji} + b_i \rho \leq \theta\ x_{0i} \qquad \text{(for all } i\text{, with } i = 1,\ldots,m\text{)}$$

$$\lambda \geq 0.$$

2.3.6 Interpretation of Homogeneous Weight Restrictions in Terms of Production Trade-offs

The CCR model without weight restrictions identifies a composite DMU that is feasible in the production possibility set. The new term added to each restriction in the envelopment form is therefore a modification of this composite DMU, and according to Podinovski (2002: 7), has the following interpretation: we assume that it is possible to change the outputs by the vector $[a_1 \quad a_2 \quad \ldots \quad a_r]$ provided that the inputs are changed by the vector $[b_1 \quad b_2 \quad \ldots \quad b_i]$, in the proportion $\rho \geq 0$.

Using this approach, we can interpret the meaning of each one of the weight restrictions presented in (2) as follows:

$v_1 - v_2 \leq 0 \Leftrightarrow$ The outputs should remain constant, if a decrease by one unit in input 1 is compensated by at least an increase of one unit in input 2.

$-2v_1 + v_2 \leq 0 \Leftrightarrow$ The outputs should remain the same if a decrease by one unit in input 2 is compensated by an increase of at least two units of input 1.

Whilst the introduction of a weight restriction implies a *reduction* in the feasible admissible region of the *multiplier form* of the model, it implies an *extension* of the feasible region of the *envelopment form* of the model. Given that the feasible region of the envelopment form represents the production possibility set, we can indeed conclude that the production possibility set of a model with weight restrictions is broader than that of a model without weight restrictions. Therefore, the value obtained for θ^* preserves its meaning as a radial reduction applicable to all the inputs. However, after the introduction of weight restrictions, the maximum radial reduction deemed feasible may in fact be greater, leading to a decrease in the value of efficiency for the DMU under evaluation.

2.4 Conclusion

In this chapter we have discussed the technique that we have used for analysis. We started by discussing the basic CCR model and then discussed the different returns to scale assumptions that can be included in DEA. Furthermore, we discussed the measurement of cost efficiency and allocative efficiency. Finally, we discussed the need to include weight restrictions, the different approaches that can be used, and the interpretation of the weight restrictions.

3 Comparing the Performance of Primary Care Providers: A Literature Review

3.1 Introduction

In this chapter, we review the methods that can be used to compare primary health care providers and the empirical studies that have been carried out. This literature review also considers the issues that still need to be addressed in order to develop a performance assessment framework that can lead to performance improvement in primary care.

We start by discussing the nature of primary care. Secondly, we review the use of Performance Indicators (PIs) and discuss some of the frameworks that have been proposed to compare primary health care providers. Thirdly, we discuss the studies that have used DEA and parametric techniques to compare primary health care providers. Finally, we discuss some of the potential unintended consequences that can result from performance assessment and distinguish between summative and formative evaluation.

3.2 The Nature of Primary Care

The recognition that primary care is a very important part of any national health care delivery system and the recognition that primary care provision is responsible for a significant proportion of the health care budget, suggest that efforts should be made to identify systematic variations in performance between primary care providers, with the view to improving performance across the system.

However, there are some considerable challenges in performing a fair comparison. Perhaps the greatest challenge in comparing primary care providers relates to the nature of primary health care provision. Whilst a hospital is an organisation with clear boundaries, where patients are admitted and discharged; primary health care delivery is an open community-based system with unclear boundaries. Furthermore, primary care has a fundamental role in terms of providing a long-term, continuing and generalist support towards individuals and families (Fry 1993). This long-term, continuous relationship is essential to build trust between the health care professionals and the patients and is one of the founding principles of primary care. In attempting to evaluate primary care providers, it is therefore important to take this aspect into account and aim to develop a framework that recognises the holistic nature of primary care.

3.3 Frameworks used to Evaluate Primary Care Providers

Several frameworks have been proposed to evaluate primary care providers. These frameworks tend to rely on the use of PIs detailing different aspects of quality in health care delivery.

Donabedian (1980) provides a framework that can be used to classify the different PIs into categories. This framework is characterised by three essential elements in health care delivery: structure, process and outcome. Structure refers to the "relatively stable characteristics of the providers of care, of the tools and resources they have at their disposal, and of the physical and organizational settings in which they work" (Donabedian 1980: 81). In this respect, it refers to what is known in the production literature as inputs. The term process refers to the activities that take place between health care practitioners and patients, and the term outcome refers to the impact on the patients' current and future health status that can be attributed to the health care provided to the patient (or to the lack of medical care). Donabedian (1980) argues that an evaluation of medical care needs to study these three elements together, because they are interconnected and work as a whole.

By focusing on different aspects of Donabedian's model, several different frameworks have been proposed to compare primary health care providers. For example, Baker (1992) proposed a list of 76 PIs related to the structure and process of General Practices in England. Eight categories of items were considered: (1) equipment, (2) clinical activities, (3) the team composition, (4) records, (5) organization, (6) premises, (7) availability of care and (8) clinics. If a particular feature was present a score of one was allocated to the practices and from the consideration of the 76 PIs, a global score was calculated and used to compare the practices in Gloucestershire, Avon and Somerset. This global score was then regressed on a number of explanatory variables to explain the variability found between the scores for different surgeries. Training practices and practices which employed a practice manager were found to be more developed. Furthermore, the lower the mean age of partners and the greater the total number of patients, the more developed were the practices. The underprivileged area score for the practices addresses was also found to explain part of the variability, with practices located in underprivileged areas presenting lower scores.

Van den Hombergh et al. (1998) also proposed a list of 129 PIs regarding the different aspects of structure and quality of management in general practice. The PIs proposed related to the following five aspects: (1) premises and equipment, (2) delegation and collaboration, (3) service and organization, (4) record keeping and (5) organisation of quality improvement programs. This framework was then used to compare a sample of 88 practices in the Netherlands. From these results, it was concluded that single-handed GP practices had less equipment and delegated fewer tasks to assistants, but scored better in terms of accessibility and availability of care. Furthermore, in agreement with the results obtained by Baker (1992), it was found that training practices scored higher in terms of equipment availability, task delegation and organization of preventive services.

These frameworks are deficient however, as although they can be used to compare primary care providers and to point towards areas that can be improved, it is also desirable to include information regarding outcomes achievement, in order to provide an indication regarding the success in making an impact in the health of the patients – the ultimate aim of primary care. The importance of measuring outcomes in health care relates with the fact that patients demand health, not health care. It is the expected benefit from diagnosis and treatment that justifies the action of going to the doctor.

Smith (1996) and Williams (1996) also emphasise the importance of measuring outcome in the public sector. They argue that, whilst for a private good, the consumer's willingness to pay for a good offers an indication of its value and impact on society, in the public sector the absence of a market means that there is a need to measure the impact of the services delivered. Smith (1996) suggests that, in the public sector, outcome measurement has both a retrospective and a prospective role. "In its retrospective role, outcome measurement can be used to determine whether the expected benefits of a public sector programme have materialized. ... In its prospective role, outcome measurement is used to guide public sector resource allocation decisions" (Smith 1996: 4, 5).

McColl et al. (1998) proposed an evidence-based approach to monitor the performance of primary care providers, which includes measures of outcome. The use of evidence based PIs is defended in order to guarantee that for each indicator there is clinical evidence that an improvement in the value of the PI *means* an improvement in the health of the patient. A list of evidence based PIs was suggested for specific clinical interventions in primary care. For example, for the treatment of hypertension, 7 PIs are suggested, such as: proportion of population whose blood pressure (BP) has been recorded in the past five years and proportion of population with a BP recording under the desirable threshold. This approach to performance evaluation could be taken as one component of the internal performance assessment framework in primary care. However, as recognised by McColl et al. (1998), primary care delivery is not solely about applying the guidelines from clinical research. There are other important dimensions of performance, which would not be adequately monitored using this type of PIs, such as patient satisfaction and efficiency. Broader frameworks have also been proposed and we shall now focus on them.

Campbell et al. (1998) have undertaken a national survey of Health Authorities (HAs) to identify PIs in use or in consideration for use in general practice. A total of 240 indicators were identified in terms of access, organizational performance, preventive care, chronic disease management, prescribing and gatekeeping. The validity of these indicators was then investigated by consulting a sample of HA managers, general practice course organizers and GPs. This type of approach aims to establish the face validity of indicators, that is, the extent to which the professionals being evaluated can agree on the appropriateness of the indicators. Indicators regarding access, organizational performance, preventive care and chronic disease management were shown to have high face validity, whilst indicators regarding prescribing and gatekeeping (appropriateness of referral decisions) were shown to have lower face validity. Campbell et al. (1998) emphasise the partiality of the PIs identified, given that no indicators had been proposed to measure effectiveness of communication, care of acute illness, health outcomes and patients' evaluation. The importance of distinguishing between indicators designed to measure the achievement of minimum standards and indicators designed to reward higher levels of performance was also highlighted.

Proctor and Campbell (1999) were also concerned with the face validity of PIs in general practice. They used semi-structured interviews with GPs, nurses and practice managers in England in order to develop a framework for performance assessment in general practice. Seven priority areas were identified: patient satisfaction; clinical activity; service development and innovation; access; health promotion; cost effectiveness and outcomes. Proctor and Campbell (1999) point out the conflicts between the views of the different professionals regarding what is good primary care, suggesting that it might be difficult to develop a framework that pleases everyone.

Greenhalgh and Eversley (1999) review the literature on the definitions of quality in general practice and also suggest that "perspectives on quality can and do conflict across different professions" (Greenhalgh and Eversley (1999: 63). In this respect, they argue for a pluralistic framework for quality assessment in primary care. It is suggested that the specific PIs used for evaluation should be locally developed and should portray the views of the different interest groups in primary care provision. An example of a framework developed at the HA level in two London districts is presented. It includes PIs covering 15 areas of primary care provision, from screening to disease management and outcomes of care. Five performance bands are proposed for each type of indicator from Band A (unacceptable performance) to Band E (leading edge) and each organisation is then assessed on how it scores on each indicator.

The importance of including measures of health outcomes in performance assessment is also visible in the UK Government's policy for the NHS. After undertaking a consultation exercise, the National Health Service Executive published a

Performance Assessment Framework (PAF) (1999), which includes 42 high level PIs to compare HAs, where 10 of the indicators relate to outcomes. Appendix A presents the list of high level PIs included in PAF. This framework includes six dimensions of performance:

(1) health improvement, aiming to reduce mortality rates and risk factors in the population;

(2) fair access, aiming to provide services in relation to need, irrespective of geography, socio-economic or demographic group;

(3) effective delivery of appropriate health care, aiming to improve clinical effectiveness and timeliness of services delivery;

(4) efficiency, aiming to reduce the costs per unit of care;

(5) patient/carer experience, aiming to improve patients' satisfaction with the services delivered;

(6) health outcomes of the NHS care, aiming to reduce the levels of disease and complications and to improve the quality of life for patients.

Despite the fact that an effort was made to develop a broad framework, covering measures of structure, process and outcome, it is not clear how this framework can be used to improve performance, because the different PIs are analysed individually and this can lead to distorted analyses. For example, Giuffrida, Gravelle and Roland (2000) investigated the variability of one of the PIs used in PAF and warn against its isolated use to assess performance. They studied the variability on hospital admission rates at the level of Family Health Service Authorities (FHSAs) between 1989/90 and 1995/6 and found confounding problems due to the multiple inputs and multiple outputs nature of primary care delivery. In this respect, they point out the importance of taking into account the level of resources used, the socio-economic environment and the level of alternative services delivered. Standardisation of the performance measures, cluster analysis, DEA and multiple regression analysis are proposed as alternative methods that can be used to perform a fairer comparison of primary care providers.

The limitations of PIs in performance assessment have also been pointed by Thanassoulis, Boussofiane and Dyson (1996). They have compared the use of PIs and DEA as alternative techniques to measure the performance of District Health Authorities in England. It was concluded that PIs are not suitable for setting performance targets for efficiency improvement, given that they only take into account part of the information. In contrast, DEA was found to be very useful for this purpose because it can take into account all the resources and all the outputs of production, as long as these inputs and outputs can be measured. The authors concluded by suggesting the combined use of these two tools: DEA should be used to measure efficiency and PIs can be used to better explain the DEA results.

The review of the different frameworks proposed has shown that considerable progress has been made in terms of the development of comprehensive frameworks for performance assessment in primary care by using PIs related to different aspects of primary care delivery. However, a framework can better contribute to performance improvement if an attempt is made to establish a link between resources usage, services delivery and outcomes achievement.

The importance of undertaking such an evaluation should not disregard the difficulties involved in this task and the potential limitations of this type of evaluation. Firstly, there are considerable difficulties in measuring outcomes in primary care and secondly, using a production metaphor can be seen as a restrictive framework for performance assessment in primary care. In the next chapter we discuss these difficulties in more detail.

Multiple regression analysis, and frontier based analysis can be used to link resources, outputs and outcomes. The relative efficiency of providers can be calculated by comparing the level of resource usage with the level of services delivered, and the relative effectiveness of the providers can be calculated by comparing the level of services delivered with the level of outcomes achieved (Gardner 1998). Most empirical studies have compared primary care providers in terms of efficiency but have failed to include measures of outcome achievement. We will now review these studies.

3.4 Beyond PIs – Using Regression and Frontier Based Analysis to Compare Primary Care Providers

We can classify the methods that can be used to measure efficiency into stochastic and deterministic, and into parametric and non-parametric. Firstly, stochastic methods include an error term, whilst deterministic methods assume that all the deviations from the model estimate are due to inefficiency. In this respect, DEA is a deterministic method. With regards to the second type of classification, DEA is a non-parametric method, because it does not impose a functional form on the production transformation of inputs to outputs. On the contrary, parametric techniques require the specification of a functional form for the production function. In terms of parametric techniques, there are two types of parametric approaches to compare the efficiency of providers: frontier based approaches, also known as Stochastic Frontiers Analysis (SFA) and regression based approaches. Frontier based approaches aim to estimate the maximum amount of outputs that should be produced from the inputs, whilst regression-based approaches estimate the average amount of outputs that can be produced from the inputs.

3.4.1 The Parametric Approach – the Use of SFA and Regression

Defelice and Bradford (1997) applied SFA to study the issue of differential efficiency between solo and group physician practices. They applied SFA to data from 924 primary care physicians in solo and group practices working for several Health Maintenance Organisations (HMOs) in the USA between 1984 and 1985. The dependent variable used was the logarithm of the number of physician visits per week. This variable was regressed on a large number of explanatory variables accounting for physical inputs and for physician effort. A dummy variable was included to distinguish between physicians working on solo and physicians working on group practices. From the results of a non-linear model, the authors concluded: "it seems that group practice is not, *ceteris paribus*, a mechanism which fosters efficiency in the production of office visits" (Defelice and Bradford 1997: 462).

Giuffrida, Gravelle and Sutton (2000) used regression to estimate a cost function for primary care administration using panel data for Family Health Service Authorities (FHSAs) in England. The variables used to explain the administrative costs of the FHSAs were the number of general practitioners, practice nurses, ophthalmic medical practitioners, dentists and community pharmacies working for each FHSA. Information regarding the number of GPs reaching the 90% target for cervical cytology screening was used as a quality proxy, and the standardized mortality rate was used to control for population characteristics. The cost function estimates were then used to compare the relative efficiencies of the FHSAs and it was observed that there was little difference in the estimated efficiency scores for the different FHSAs. Furthermore, caution regarding the use of cost estimates to label certain FHSAs as inefficient was emphasized given the potential endogeneity of some of the explanatory variables. In this respect, the authors called for careful specification of the decision-making process that generates health care costs.

Bryce, Engberg and Wholey (2000) examined the efficiency of 585 HMOs using three alternative techniques: DEA, SFA and fixed effects regression. The results suggested that the techniques provide similar results in terms of the general trend of efficiency evolution for the industry. However, there were differences regarding the individual estimates of performance, and regarding which DMUs were rated as efficient by each method.

Giuffrida and Gravelle (2001) also compared DEA and SFA estimates of efficiency in primary care. They were interested in investigating the robustness of the efficiency results between alternative models and methods. From the results, the authors concluded "DEA average efficiencies are similar to those from the regression based methods" (Giuffrida and Gravelle 2001: 170). As for the robustness of the results for individual DMUs, the authors concluded that under alternative models, DEA results appear to be robust. The same applied to alternative econometric models. However, correlations between regression and DEA individual efficiencies were low, suggesting that the "choice between DEA and regression does make a considerable difference to the rankings" (Giuffrida and Gravelle 2001: 171).

It can be argued that SFA has a fundamental advantage over DEA; it includes a noise term, accounting for measurement errors and other non-systematic disturbances in the data. In standard DEA no allowance is made to this possibility; the deviation from the frontier is taken to be totally due to inefficiency[2]. Another advantage of SFA is that the

---

[2] Some authors have proposed stochastic versions of DEA. Sengupta (1996) provides a method to estimate the efficiency distribution approach in DEA, allowing for the inclusion of a noise term.

problem of variable selection can be addressed by running statistical tests of significance. However, SFA's advantages come at a price; the potential for specification bias, given that a parametric form for the production function has to be chosen. In contexts such as public service delivery where the transformation function is difficult to specify, it is argued that the non-parametric approach is more appropriate (Charnes, Copper and Rhodes 1978). We will now review the studies that have used DEA to compare primary care providers.

3.4.2 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 Whynes 1996, Salinas-Jiménez and Smith 1996, Giuffrida 1999, Buck 2000 and Giuffrida and Gravelle 2001). Pina and Torres (1992), Gõni (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 HMOs (Rosenman, Siddarthan and Ahern 1997,

---

Following a different approach, Olesen and Petersen (1995) use chance constrained efficiency (CCE) evaluation to allow for the inclusion of endogenous perturbations in the input-output set. Efficiency scores are calculated allowing for all perturbations within a specified confidence region. CCE evaluation requires panel data to calculate the confidence regions.

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 thesis, as our focus is on primary care evaluation.

Table 3.1 (at the end of this chapter) presents a summary of the studies that have used DEA to compare the efficiency of primary health care providers and we will now review some of these studies in more detail.

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. For example, output measures included the number of immunizations and the number of cervical cytology claims.

Similarly, effectiveness was defined as the achievement of Government targets for cytology and child immunization. When we consider that preventive care represents only part of the functions of primary care, the study is potentially biased and there is a need to include measures of service delivery and outcomes achievement in terms of the other functions of primary care.

Szczepura et al. (1993) concluded that the efficiency ratings did not appear to vary according to practice size for surgeries with a single location. However, for surgeries with multiple locations, larger surgeries were found to be on average more efficient. It also did not appear that efficiency related significantly to the practices' characteristics. As for the relationship with effectiveness, the study showed the difficulty of some efficient practices in achieving the effectiveness target levels without further resources.

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, Boussofiane 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. In particular, the authors were concerned with the inclusion of output quality information in the models. Measures of activity performed were included as the main outputs. In addition, they proposed the inclusion of satisfaction measures in order to reflect the quality of the services delivered. Furthermore, an attempt was made to include data regarding the quality of the outcomes. They included the number of babies at risk as an input and then included the number of survivals from babies at risk as an outcome. Weight restrictions were used to link the weights of these two variables.

Another extension proposed in Thanassoulis, Boussofiane and Dyson (1995) was a model based on the specification of ideal input-output levels, from the decision-makers point of view. This model aimed at determining targets as close as possible to these levels, allowing the decision-maker to set its own priorities regarding performance improvements. The study's emphasis was on enhancing the DEA models towards the achievement of more useful results for management.

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. Preferred practice patterns are identified using a two-stage DEA model. In the first stage, a standard DEA model without weight restrictions is run to identify the best practice frontier and the optimal weight bounds. In the second stage, these weight bounds are imposed in the form of assurance regions, and the DEA model is re-run to determine the restricted best practice group of physicians. 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-Jimenéz 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. The argument behind this approach is that having a good structure of care and following sound processes of care increases the probability of producing good outcomes. More recently, Davies and Crombie (1997); Crombie and Davies (1998) have also defended the use of process measures in the measurement of performance. 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.

García et al. (1999) also wanted to analyze the sensitivity of the results to different sets of outputs. The results suggested a significant amount of sensitivity to model specification, pointing towards the importance of model validation. This is particularly important because performance assessment in primary care is necessarily partial, as we cannot measure all the essential quality aspects.

Our review of the literature so far points towards the fact that most DEA studies applied to primary care have failed to include measures of outcome achievement, when this measure is an important part of the evaluation in primary care. Having reviewed the studies that have used DEA to compare primary care providers, we will now discuss some of the unintended consequences that can result from performance assessment.

3.5 Moving Towards Formative Evaluation

Smith (1995) pointed out that performance assessment in the public sector can lead to dysfunctional behaviour. He identified five potential unintended consequences that can result from the publication of performance measures:

(1) tunnel vision, which relates to the focus on the areas being measured at expense of other important aspects;

(2) sub-optimisation, which can occur when managers focus on narrow local objectives, and ignore the broader organizational objectives;

(3) myopia, when a focus on short-term measures takes prevalence over long-term issues;

(4) misrepresentation, which relates to the potential for altering the data in order to show better performance results;

(5) gaming, which relates to the potential for altering the behavior in a dysfunctional way, in order to achieve better performance results.

Casalino (1999) also emphasized the potential for dysfunctional responsive behaviour to performance assessment in health care. Taking account of these potential problems, Smith (1995) and Goddard, Mannion and Smith (2000) outlined a number of strategies that can help addressing these potential problems. One of the strategies suggested relates to the need to involve staff at all levels when developing a performance assessment framework, and the need to move beyond the use of performance measures for control. This relates to an important aspect regarding performance assessment; the perceived purpose of the assessment exercise and the way the results are interpreted. Dyson (2000) points out situations in which the use of league tables and rankings can misrepresent and exacerbate the situation under evaluation and suggests the use of DEA as an exploratory tool to promote learning.

In this respect, it is our conviction that performance assessment in primary care can be most useful if the perceived purpose is one of learning and improvement, and as long as the results are interpreted very cautiously, in a formative rather than judgemental way. We feel that the theory behind program evaluation, with its emphasis on providing useful information regarding ways to improve programmes and interventions, can offer some useful guidelines in performance assessment.

Patton (1986: 14) defines program evaluation as "the systematic collection of information about the activities, characteristics, and the outcomes of programs for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs are doing and affecting". Despite the apparent similarities between performance assessment and program evaluation, these two research areas have been concerned with different questions, and have tended to use different methods.

Scriven (1967) distinguishes between two types of evaluation, depending on its objectives and on the types of methods used, summative evaluation and formative evaluation. Summative evaluation involves the collection and analysis of data with the aim of determining whether a program should be continued or stopped and whether an organisation is performing better or worse than another. Its aim is solely descriptive and classificatory. In contrast, formative evaluation involves the collection and analysis of data with the aim of developing and improving the program or the organisations. Formative evaluation studies are characterized by an effort to involve the stakeholders and to provide useful feedback to the decision-makers and policy makers in terms of organizational improvement and knowledge development. The ultimate aim of a formative evaluation study is the promotion of learning.

The integration between program evaluation and performance assessment literatures and methods can be very valuable, given that each one of these traditions has been concerned with different aspects. While quantitative performance assessment studies have been mostly concerned with the comparison of large numbers of organisations aiming at description and classification, evaluation studies have been mostly concerned with the in-depth analysis of programs and their success or failure in different contexts.

Blalock (1999) also calls for the integration between these two research areas: "used as *complementary* tools, the two movements can offer more valid and reliable information to decision makers, and can more accurately guide improvements in programs" (Blalock 1999: 142). It is this type of integration that we aim to achieve in this thesis. In order to obtain results that can be used in practice, we have developed the framework for performance assessment with a sample of PCG/Ts, as well as validated and discussed the results with these PCG/Ts and the participating GP surgeries.

3.6 Conclusion

In this chapter we have discussed the nature of primary care and then reviewed the literature on the evaluation of primary care providers. The use of PIs is discussed and it is concluded that they are inadequate in themselves, as they typically do not link resources to outputs and outcomes. We also reviewed the use of parametric methods to compare primary care providers. The fact that these methods require the introduction of assumptions regarding the functional form of the transformation function was pointed out as a drawback, because in primary care the technology involved in transforming inputs into outputs is not well understood.

DEA is then introduced as a performance evaluation methodology that can include and link multiple inputs and outputs and therefore in principle encompass structure, outputs and outcomes. DEA studies of primary care are then reviewed and the various approaches to defining the inputs and outputs required for a DEA evaluation considered. It is shown that most of the previous studies have 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 chapter a model is developed for holistic, inclusive and formative evaluation of primary care.

Table 3.1: Summary of the DEA applications to primary health care

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Bates, Baines and Whynes (1996)	106 General Practices in the Lincolnshire HA, for the year of 1993/94.	Level of prescribing expenditure	Number of patients treated. Patients grouped by age category, exempted from payment for prescriptions, and temporary residents.	None.
Bryce, Engberg and Wholey (2000)	A total of 585 HMOs practicing in USA. However, the sample size per year varied from 151 to 333. Data refers to 1985 - 94	(1) Hospital days; (2) Ambulatory visits; (3) Administrative expenditure; (4) Other expenditures.	Total member-years of coverage provided by the HMO during the year.	None.
Buck (2000)	English Community Dental Services by Health Authority. Data refers to 1997/98.	Hours worked by dental officers, therapists, hygienists and others.	Number of screening episodes, number of treatment episodes, and number of preventive episodes.	None.
Chilingerian and Sherman (1996)	326 Primary Care Physicians (PCPs) working in HMOs in USA. Data refers to 1990.	(1) Quantity of medical-surgical office visits. (2) Quantity of ambulatory surgery procedures. (3) Quantity of medical-surgical hospital days. (4) Quantity of mental health visits. (5) Quantity of emergency room visits. (6) Quantity of radiology, lab and diagnostic tests. (7) Quantity of therapy visits. (8) Quantity of referrals resulting in a specialist visit.	(1) Quantity of male and female enrolees aged 0-19. (2) Quantity of female enrolees aged 20-39. (3) Quantity of male enrolees aged 20-39. (4) Quantity of female enrolees aged 40-59. (5) Quantity of male enrolees aged 40-59. (6) Quantity of female enrolees aged 60+. (7) Quantity of male enrolees aged 60+.	None. Specialists and generalists are supposed to face the same production frontier and therefore, are compared with each other.

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Chilingerian and Sherman (1997)	326 Primary Care Physicians (PCPs) working in HMOs in USA. Data refers to 1990.	(1) Quantity of medical-surgical office visits. (2) Quantity of ambulatory surgery procedures. (3) Quantity of medical-surgical hospital days. (4) Quantity of mental health visits. (5) Quantity of emergency room visits. (6) Quantity of radiology, lab and diagnostic tests. (7) Quantity of therapy visits. (8) Quantity of referrals resulting in a visit by a specialist.	(1) Quantity of male and female enrolees aged 0-19. (2) Quantity of female enrolees aged 20-39. (3) Quantity of male enrolees aged 20-39. (4) Quantity of female enrolees aged 40-59. (5) Quantity of male enrolees aged 40-59. (6) Quantity of female enrolees aged 60+. (7) Quantity of male enrolees aged 60+.	None. Specialists and generalists are supposed to face the same production frontier and therefore, are compared with each other.
García, Marcuello, Serrano and Urbina (1999)	54 Spanish Primary Care Centres. Data refers to 1996.	(1) Total number of General Practitioners (GPs). (2) Total number of nurses. (3) Number of lab prescriptions ordered by the center. (4) Number of X Rays ordered by the center. (5) Number of referrals to specialist doctors. (6) Prescriptions costs for active population and their families. (7) Prescription costs for old age pensioners and their families.	(1) Total number of consultations carried out by the professional working at the Centre. (2) Number of consultations made by every individual of the zone covered by the Centre. (3) Synthetic index that includes coverage of the programs implemented by the Centre, weighed by technical complexity. (4) Compliance with the Minimal Technical Standards (MTS) of quality in the performance of the services.	None. However, we should note that some degree of standardization based on the characteristics of the population and on case mix is already included in the output variable (3).

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Giuffrida (1999)	90 Family Health Services Authorities (FHSA) in England. Malmquist indexes calculated from 1990/91 to 1994/5.	(1) Number of unrestricted GPs in the FHSA. (2) Number of practice nurses (WTE) in the FHSA.	In the full model: 5 output variables accounting for the number of patients registered with the FHSA, accounting for deprivation areas, and classified into age classes. 5 output variables accounting for GPs achieving high levels of immunisation; pre-school boaster; cervical cytology; health surveillance to children. 2 quality proxies, accounting for GP's list size and for practices characteristics. The reduced model only includes 4 outputs and the 2 quality proxies.	(1) Number of deaths from all causes of FHSA residents aged 0-64. (2) Area of the FHSA in hectares.
Giuffrida and Gravelle (2001)	90 FHSA in England. Data refers to 1993-4 and 1994-5.	Model 1: As in Salinas-Jiménez and Smith (1996). Model 2: As in Salinas-Jiménez and Smith (1996). Model 3: (1) Synthetic Cost: Logarithm of the remuneration of GPs and practice nurses in the FHSA. (2) Logarithm of the number of GPs for whom the FHSA is the responsible committee.	Model 1: As in Salinas-Jiménez and Smith (1996). Model 2: As in Salinas-Jiménez and Smith (1996), but excluding two variables: Outputs (3) and (4).	Models 1 and 2: Standardized Mortality Rate (SMR) from all causes of FHSA residents aged 0-64. Model 3: (1) SMR; (2) Area of the FHSA; (3) Number of patients classified as deprived.

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Göni (2000)	31 primary care teams in Spain Data refers to	(1) Medical and non-medical personnel; (2) Pharmaceutical costs; (3) Laboratory costs.	(1) Number of patient initiated consultations; (2) Number of programmed consultations; (3) Time devoted per consultation per professional.	None.
Huang and McLaughlin (1989)	77 primary care clinics in America. Data refers to 1978-1983.	(1) Whole Time Equivalent (WTE) Administrative; (2) WTE Medical Technician; (3) WTE Physician; (4) WTE New Health Professional (5) WTE nurse	(1) Number of Physician encounters (2) Number of Non Health Professional encounters (3) Number of Nurse encounters	Inputs: (1) Population size; (2) Program age; (3) User's age. Output: Total number of encounters
Luoma et al. (1996)	202 Finish Health Centres Data refers to 1991.	Operating costs excluding rehabilitation costs and costs of purchased services.	Outpatient care: (1) Health care and medical care visits to a physician; (2) Health care visits to other personnel; (3) Medical care visits to other personnel; (4) Visits of supervised domiciliary care; (5) Dental care visits; (6) Special examinations. Inpatient care: (1) Short-term inpatient days; (2) Long-term inpatient days for heavily dependent patients; (3) Long-term care for other patients.	None.
Ozcan (1998)	160 physicians treating otitis media in the state of Virginia, in USA. Data refers to 1993.	(1) Primary care physician visits. (2) Specialist visits. (3) Inpatient. (4) Prescriptions. (5) Laboratory procedures.	(1) Low severity patients; (2) Medium severity patients; (3) High severity patients.	Clustering based on rural versus metropolitan based physicians and clustering based on four regions of Virginia (Central, North, Southeast and West).

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Pina and Torres (1992)	10 Health Centres from the Province of Huesca, Spain. No reference is made to the year to which the data refers.	(1) Personnel costs; (2) Medication Costs; (3) Other costs.	(1) Average number of consultations per inhabitant; (2) Average number of consultations per professional per working day; (3) Percentage of programmed consultations.	In one of the models, Personnel costs were considered as non-controllable.
Rosenman, Siddharthan and Ahern (1997)	28 Health Maintenance Organizations (HMO) licensed to practice in the state of Florida. Data refers to 1994.	(1) Total assets. (2) Total administrative expenses. (3) Total medical care expenses	Total number of enrolees into three categories: Medicare, Medicaid and Commercial.	None.
Salinas-Jeminéz and Smith (1996)	90 FHSA in England. Data refers to 1990/91.	Gross expenditure on General Medical Services per head of resident population.	(1) General medical practitioners (GPs) per 10,000 patients on lists. (2) The % of practices employing a practice nurse. (3) The % of GPs who had a patient list of less than 2,500 patients. (4) The % of GPs not practising single-handed. (5) The % of GPs who had achieved the higher rate of payments for childhood immunization. (6) The % of females aged 35 to 64, registered with the FHSA and who had an adequate cervical smear in the previous five and a half years. (7) The % of practice premises which satisfied the minimum standards.	(1) Standardized illness ratio. (2) Unemployment.

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Schinnar et al. (1990)	54 mental health partial care programs New Jersey, in USA. Data refers to 1984/85.	Efficiency: (1) Salaries and benefits to all care personnel (2) All other costs Effectiveness: Partial care contacts per client	Efficiency: (1) Number of active clients (2) Partial care days provided per client Effectiveness: (1) Percent of "okay" discharges (2) Average percent of mental health needs met per client (3) Average number of mental health needs met per client (4) Average percent of social needs met per client (5) Average number of social needs met per client	
Siddharthan, Ahern and Rosenman (2000)	164 HMOs practising in USA. Data refers to 1994.	(1) Inpatient days for adults. (2) Total number of inpatient days for maternity care broken down by stays for mothers and newborns and tests performed in an inpatient setting. (3) Outpatient visits. (4) Emergency room visits. (5) Procedures performed at the ambulatory surgery centre of outpatient clinic.	Total number of enrolees into three categories: Medicare, Medicaid and Commercial.	None.
Szczepura, Davies, Fletcher and Boussofiane (1993)	52 Practices from three health districts in Warwickshire, England. Data refers to the period of 1989 to 1991.	(1) WTE general practitioners; (2) WTE practice nurses (3) WTE administrative staff (sum of WTE practice manager, receptionist, and secretary)	(1) Total number of immunizations (2) Total number of child development screens (3) Total number of cervical cytology claims (4) Total number of contraceptive claims (5) Weighted number of patients aged 65-74 and 75+	Clustering according to the practice with one surgery or practices with multiple surgeries.

Study	Level of Analysis (DMUs)	Inputs	Outputs	Non-controllable Variables
Thanassoulis, Boussofiane and Dyson (1996)	83 District Health Authorities in England, which provided perinatal care.	(1) Whole Time Equivalent (WTE) Obstetricians; (2) WTE Paediatricians; (3) WTE midwives; (4) WTE nurses; (5) Number of babies at risk.	(1) Number of birth episodes performed; (2) Number of deliveries to mothers resident in the area; (3) Number of special care consultant episodes; (4) Number of intensive care consultant episodes; (5) Number of abortions; (6) Number of babies at risk surviving; (7) Number of very satisfies mothers; (8) Number of satisfied mothers.	Outputs 1, 2, 3, 4 and 5. Input 5.
Zavras et al. (2002)	133 Primary Care Centres (PCCs) in Greece. Output data refers to 1998 and inputs data refers to 1999.	Personnel by category: (1) Medical, nursing, paramedical and administrative. (2) Number of patients covered by category: Actively employed, pensioners, eligible family members.	Model 1 for all 133 PCCs: Annual number of patient visits to each PCC. Model 2 for a sample of 76 PCCs: (1) Annual number of patient visits to each PCC; (2) Annual the number of laboratory tests undertaken by each PCC.	The analyses were performed separately for four PCCs sizes: I – PCCs covering up to 10,000 patients. II – PCCs covering 10,001-25,000 patients. III – PCCs covering 25,001-50,000 patients. IV – PCCs covering more than 50,000 patients.

4 Towards a Comprehensive Framework for a Formative Evaluation in Primary Care

4.1 Introduction

In this chapter we discuss the development of a 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.

The issues and difficulties involved in developing a comprehensive framework for primary care are discussed with particular emphasis on the difficulties associated with measuring the outcomes of primary care. The eventual decision to focus on a speciality rather than evaluate primary care as a whole is discussed, together with the limitations of this approach.

4.2 The Participating PCG/Ts

Our main research objective is to undertake a formative evaluation in primary care and, for this reason; we believe it was essential to involve the decision-makers from the very beginning. In this respect, we decided to approach a number of PCG/Ts to assess their interest in taking part in this research project.

In May 2000, we wrote to the Chief Executives (CEs) of 16 PCG/Ts in Warwickshire, Herefordshire, Bedfordshire, Cambridgeshire and Suffolk. The initial correspondence provided a summary of the research objectives and suggested a meeting between myself and the CE of each PCG/T in order to discuss their interest in taking part in the research project. Rugby PCG, Bedford PCT, West-Suffolk Borders PCG, Hereford PCT, Luton PCG and South Peterborough PCT expressed an interest and separate meetings with the CE of each of these PCG/Ts took place to assess their availability to take part in this research project. After two meetings with the CEs of Luton PCG and Hereford PCT, they decided not to take part in the project due to time constraints regarding participation in the workshops.

The second stage of the research involved two rounds of workshops with a group of 6-8 members of each PCG/T Board. All the members from the Board of each PCG/T were invited to take part in the workshops, but due to timetable restrictions, it was only possible to meet with some of the Boards' members. The composition of the groups who took part in the workshops was as follows: two to four GPs, one nurse, one Lay member, the Director of finance and performance and the Board's CE. These workshops took place during the year of 2001 and lasted approximately ninety minutes each. During the first round of workshops, we discussed the research objectives, the methodology to be used and the concepts related to performance assessment in primary care.

Using the ideas generated in the first round of workshops, a simple conceptual framework was developed. This framework was the basis for the discussions, which took place during the second round of workshops. The final conceptual framework developed resulted from an attempt to include all the issues highlighted during the workshops.

4.3 The Level of Analysis

The identification of the appropriate level of analysis is a crucial part of any performance assessment exercise. With the present accountability structure in the NHS, two meaningful levels of analysis were identified. The highest level is the PCT. At the PCT level strategic decisions are taken and the resources are allocated into their several components. A comparison of several PCTs could provide useful information regarding fair access to services, efficiency and effectiveness of different strategies. Below the PCT level, there are the GP surgeries, which deliver the services to the patients and make operational decisions regarding day-to-day resource allocation. A comparison of several GP surgeries should lead to an important contribution towards gaining an insight of why there are differences in performance at the PCT level. Furthermore, a comparison of the GP surgeries within a PCT can provide useful information regarding fairness of access to services, efficiency and effectiveness, and coordination of different structures and practices. Performance assessment at this level can have an important role by supporting a process of accountability from the surgeries to the PCT.

The initial research objective was to compare all PCG/Ts in England and then focus on the performance of the four participating PCG/Ts and compare all their GP surgeries. In early 2000, contacts were established with the National Primary Care Research and Development Centre in Manchester. This group has been developing a national database for PCG/Ts, which aims to provide linked information on population characteristics, primary care provision and health status for all PCG/Ts in England. Despite initial promises regarding availability of data for resources usage, services delivered and outcomes, by September 2000, we were told that there was a significant delay in the development of the database and that only data in terms of resources usage would be available in the near future. As a result, the focus of this study became the comparison of GP surgeries in terms of their service delivery. Below we present the conceptual framework that was developed to compare GP surgeries in England and discuss the methodology used for its measurement.

4.4 Developing a Conceptual Framework for Performance Assessment in Primary Care

Wholey (1996: 146) states that performance "is not an objective reality out there waiting to be measured and evaluated". But more intrinsically, performance is socially constructed and means different things to different people. Whilst the Treasury may be concerned with budgets and financial performance; the PCT may be concerned with an efficient and equitable use of resources; the health care professionals may be concerned with meeting the individual needs of the patients, and the patients themselves may be concerned with the quality of the services they get. In this respect, whilst assessing performance appears to be associated with sorting out the good from the bad (Pollitt 1986), what is good from one point of view may actually be bad from another. For example, whilst health care professionals and patients see the amount of time spent with a group of patients as something that should be maximized, the PCT may see this as a resource that should be minimized, because it represents an opportunity cost in terms of the delivery of services to other patients. In this respect, in order to measure performance, we first need to make explicit from which perspective it will be measured.

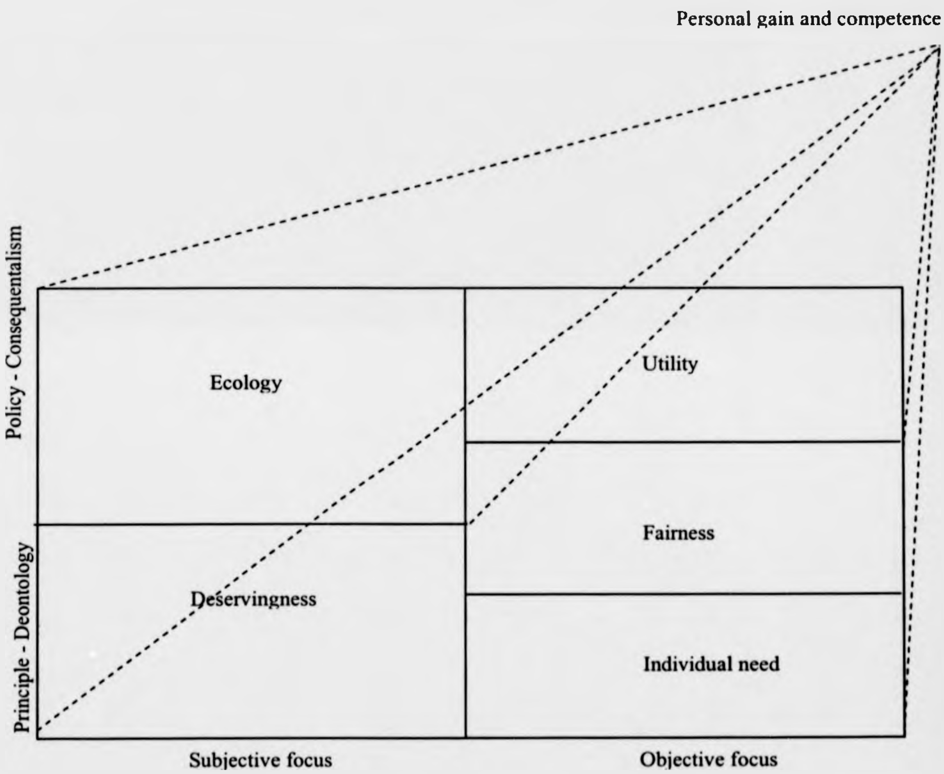
In terms of public services provision, performance assessment is legitimate from the perspectives of the several stakeholders of the services. Stakeholders are the groups of people who have a legitimate interest in the services (Smith 1996). Williams (1996) identified a long list of potential stakeholders for the NHS, such as the patients and their representatives, the taxpayers, the professionals working in the NHS, local government bodies and the Department of Health. Each one of these groups of people has its own set of objectives for the NHS and these should form the basis for the development of different performance assessment frameworks. Fitzgerald and Storbeck (2002) point out the importance of addressing the question of *whose interests* are organisations answering to, and we see this as a very important part of any performance assessment exercise.

In this study, we are concerned with performance assessment from the perspective of the PCG/T responding to external forces, in particular, Government pressure to improve primary care delivery for the local population using the limited resources available. However, it is important to recognize that a PCG/T is a very complex organization in itself and different professionals within each PCG/T will hold different objectives in terms primary care delivery, dependent upon their roles within the organization. For example, some of the GPs that belong to the PCT's Board will have concerns for the efficient allocation of resources, but also will have concerns for meeting the individual needs of their patients. In this respect, within each PCG/T there are different and potentially conflicting values at the decision-making level concerning resource allocation and service delivery. We will now discuss the different values that form the basis of decision-making in public service delivery.

4.4.1 The Role of Value Heuristics in Decision-making and its Effect on Performance Assessment

Fisher (1998) identified six value heuristics used in decision-making in the public sector: (1) individual need; (2) deservingness; (3) fairness; (4) utility; (5) ecology; and (6) personal gain and competence. Fisher (1998) discusses these value heuristics in terms of two dichotomies: Policy versus Principle and Objective versus Subjective. Figure 4.1 presents the value heuristics in a matrix using these two dichotomies.

Figure 4.1: The value heuristics of resource allocation (Fisher 1998: 137)



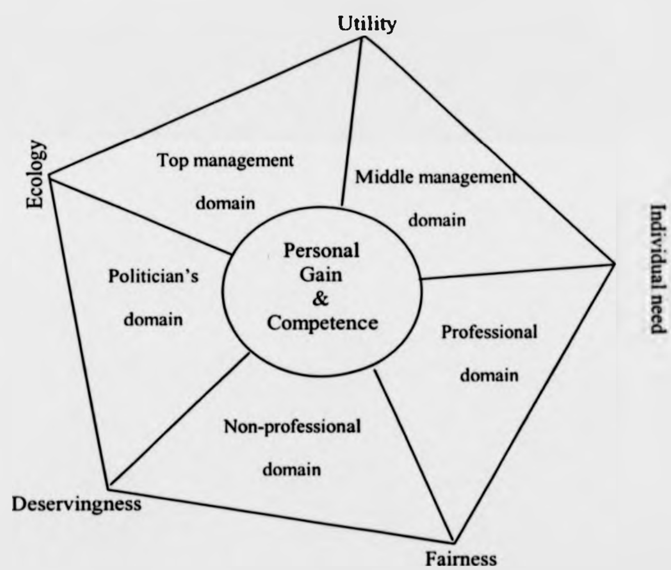
Ecology refers to the need to accommodate the views of the most influential interest groups and lobbies. Deservingness refers to the allocation of resources and services according to people's moral worthiness. According to this heuristic a patient who systematically fails to cooperate with the treatment is seen as less deserving than someone who is seen to cooperate and comply. These two heuristics reflect subjective ways of allocating resources.

From an objective standpoint, Fisher (1998) identified three heuristics that can be used in resource allocation and service delivery. Utility uses economic evaluations and cost-benefit analysis to determine the best way to allocate resources and services. According to this heuristic, the objective is to achieve the greatest good for all. Utility is concerned with the consequences of the actions and therefore, it is located in the upper side of the matrix. Fairness aims to treat everyone equally. According to this heuristic, drawing lots or a 'first come, first serve basis' are the best ways to allocate services. Fairness is concerned with the consequences of the actions, but it is also a question of principle. The heuristic of individual need is concerned with the satisfaction of the needs of each patient, independently of the consequences for others and independently of its cost-effectiveness.

Lastly, there is the heuristic of personal gain and competence, which cuts across each of the dichotomies. The application of this heuristic in service delivery means that the decisions are made in a way that *benefits* the decision-maker. The argument is that everyone uses this heuristic to a greater or lesser extent. People will try to make a decision using the appropriate methods because this leads to greater self-esteem and professional pride. People may also make the decision from which they expect to gain some personal advantage, such as keeping their job.

As Fisher (1998) points out, people do not just use one heuristic in order to make and justify their decisions. He suggests that depending on their hierarchical position in the organisation, people will use different heuristics. Furthermore, Fisher (1998) suggests that under pressure, people tend to use a different set of heuristics. In this respect, he argues that each professional faces a conflict between two main heuristics due to the contradictions associated with his or her role within the organisation. Figure 4.2 presents Fisher's model of the location of the value heuristics in public sector organisations. For example, doctors and nurses follow their training and tend to use the individual need heuristic. However, in a situation of tight budgets, for example, they use the fairness heuristic as a way of allocating resources and delivering services.

Figure 4.2: The organisational location of the value heuristics in the public sector organisations (Fisher 1998: 145)



According to Fisher (1998), each professional will take a decision that *fits* the purpose, whilst being compatible with his/her values and beliefs. When decisions are made using a heuristic process, resources are not allocated optimally. Instead, a process of argument tends to occur and the final solution is an ill-defined compromise.

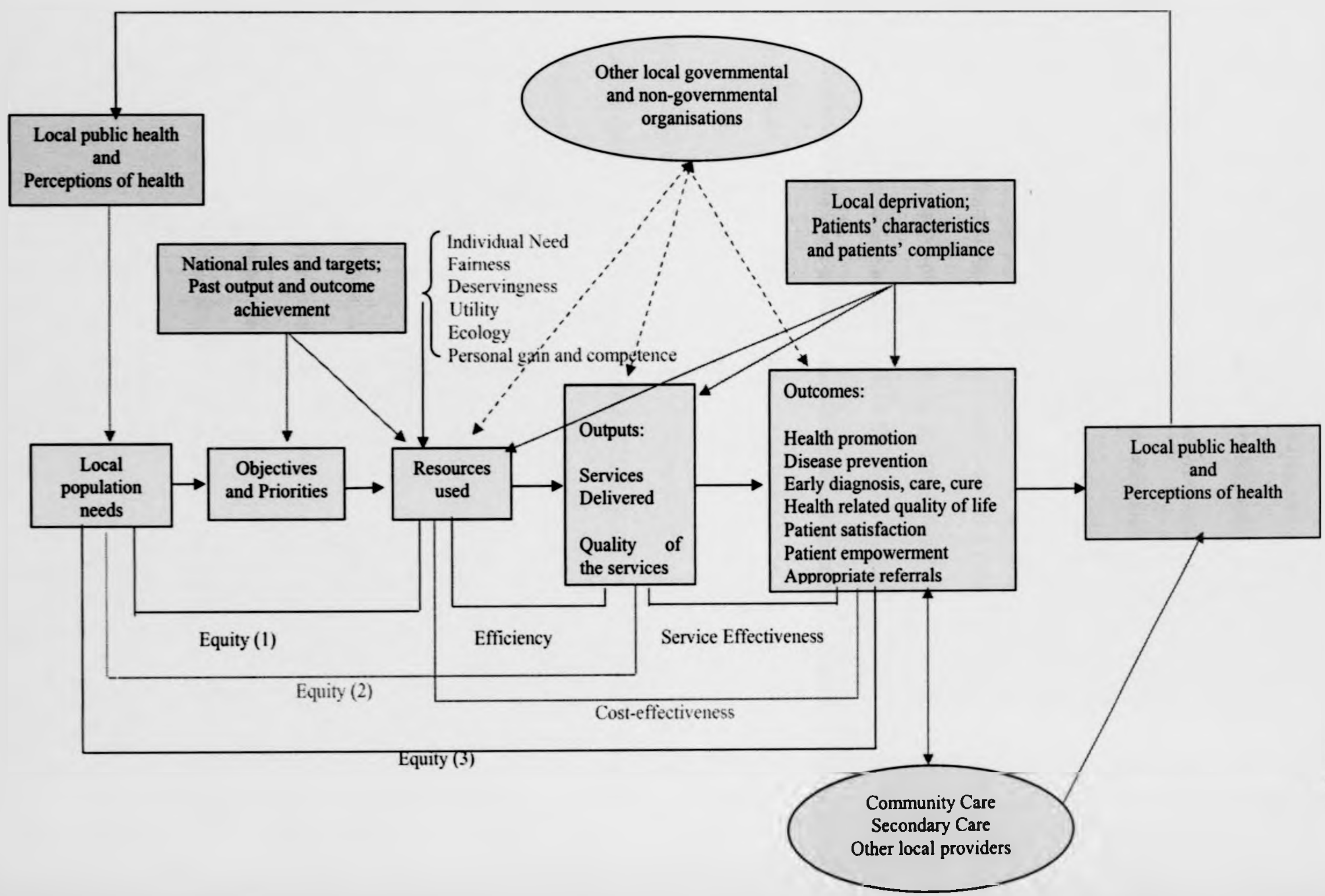
Once we recognise that public service provision is informed by multiple, incommensurable and often conflicting values, we can no longer arrive at a single view of performance. The evaluation of a public organisation has to portray different views, based on the multiple values that underlie public service provision. We will now present the conceptual framework developed for formative evaluation in primary care and discuss the different criteria used for evaluation.

4.4.2 Towards a Simplified Framework for Formative Evaluation in Primary Care

Our purpose of undertaking a performance assessment in primary care is 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 (Figure 4.3).

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.

Figure 4.3: A conceptual framework for formative evaluation in primary care at the level of GP surgeries



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.

As previously discussed, 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.

4.4.3 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.

4.4.3.1 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 2000a). 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.

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 allocativelly 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:

$$\textit{Technical Efficiency} = \frac{\textit{Weighted Sum of Outputs}}{\textit{Weighted Sum of Inputs}}$$

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

4.4.3.2 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:

$$Service\ Effectiveness = \frac{Weighted\ Sum\ of\ Outcomes}{Weighted\ Sum\ of\ Outputs}$$

$$Cost\ Effectiveness = \frac{Weighted\ Sum\ of\ Outcomes}{Weighted\ Sum\ of\ Inputs}$$

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.

4.4.3.3 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 'equalness'? 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 thesis we have concentrated on assessing horizontal equity only. We will now clarify what we mean by *equal need* and *equal treatment*.

4.4.3.3.1 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 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 'equalness'? 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 thesis we have concentrated on assessing horizontal equity only. We will now clarify what we mean by *equal need* and *equal treatment*.

4.4.3.3.1 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.

4.4.3.3.2 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.

4.4.4 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 thesis. 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).

4.5 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.

4.6 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-Jeminez 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-Jeminez and Smith 1996). We will next discuss alternative measures of health outcomes that can be used to compare primary care providers.

4.6.1 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.

4.6.2 The Need to Focus on a Specific Disease

The feasibility of extracting data from the GP surgeries of the participating PCG/Ts 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 (CHD), 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 evaluation of primary care as a whole was thus dropped at this stage, but the issue of general versus speciality evaluation is discussed later in this chapter.

Three specialities that could benefit from a formative evaluation were identified: (1) Cancer, (2) CHD and (3) diabetes. From these three specialities, we decided to focus on diabetes care for three reasons. Firstly, because of its rising prevalence and its rising financial burden to the NHS; Secondly, because there is evidence of great variations in the standards of diabetes care provided (The Audit Commission 2000b); Finally, because diabetes is a condition coordinated in primary care, the area we want to investigate.

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 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 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.

4.7 Conclusion

In this chapter we have presented the conceptual framework developed for formative evaluation in primary care at the level of GP surgeries. 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 then proposed as a methodology that can be used to measure efficiency, service effectiveness and cost effectiveness.

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 diabetes care delivery is justified in order to better establish a link between resources usage, services delivery and outcomes achievement. Finally, the advantages and disadvantages of evaluating GP surgeries by focusing on a specialty were discussed.

In the next chapter we discuss the methodology followed to develop the specific measures for diabetes care evaluation and the process followed to collect the necessary data.

5 Research Design and Methodology

5.1 Introduction

This chapter provides the background for the empirical study undertaken in this thesis. Firstly, we discuss the workshops undertaken with the PCG/Ts in order to develop the models for formative evaluation in primary diabetes care, together with the selection of the participating surgeries. Secondly, we discuss the methodology followed in this research. We have used a combination of quantitative and qualitative methods in order to better understand the context of study and to better understand the results of the comparative evaluation. We finish with a discussion of the methods used to collect the necessary data, including a discussion of both two surveys undertaken.

5.2 Research Design

In the previous chapter we have discussed why we have decided to focus on a specific disease rather than an evaluation of primary care as a whole. Once it was decided to focus on diabetes, a second round of two workshops with each PCG/T was arranged to develop the framework for performance assessment in primary diabetes care. The final framework agreed by the participating PCG/Ts is discussed in detail in chapter 6. This framework is a compromise of several drafts developed in collaboration with each PCG/T. In general, we did not find a great deal of disagreement. The definition of the clinical outcomes of care was an aspect that caused some disagreement, particularly in terms of the thresholds for some of the measures. We also found some disagreement regarding the use of measures of outcomes achievement (such as the proportion of patients with cholesterol levels under control) and measures of outcomes improvement (such as the change in the proportion of patients with cholesterol levels under control). As a result, it was decided that as part of this first exercise we should focus on outcomes achievement, as this measure would indicate how well each surgery is performing in terms of achieving the outcome targets outlined by the Government. In general, whenever disagreements were raised, research evidence was used to reach a consensus between the several members involved.

Once the final framework for evaluation was approved by all the PCG/Ts, several meetings with the information facilitators of each PCG/T were arranged in order to confirm that the necessary data could be extracted from the surgeries' records. Unfortunately, at this stage, it was found that West Suffolk Borders PCG and Rugby PCG would not be able to extract the necessary data from their surgeries' records. It was discovered that in most of their surgeries, some of the necessary data had not been collected. As a result, West Suffolk Borders PCG decided to withdraw from the research project.

However, Rugby PCG decided to undertake a process of prospective data collection over a six-month period, in order to perform a small-scale comparison between its 13 surgeries. Unfortunately, this proved to be a rather complex and unrewarding exercise as only four surgeries collected the data according to the required methodology. The data collected from these four surgeries was not used in this study, but the exercise was useful as a learning process regarding some of the complexities involved in assessing performance in primary care.

At the end of this process, there were only two remaining PCTs: South Peterborough PCT and Bedford PCT. During a workshop with Bedford PCT, it was suggested that contacting other PCTs in the region might increase the number of participating PCTs. From these contacts, several positive replies were received. However, after meetings with the information facilitators, we concluded that only Huntingdon PCT would be able to extract the necessary data. Huntingdon PCT agreed to the methodology and the framework previously developed, and became the third PCT taking part in this research project.

South Peterborough, Bedford and Huntingdon PCTs were able to extract the necessary data because they were engaged in a data improvement project called Primary Care Information System (PRIMIS). PRIMIS was set up in April 2000 by the NHS Information Authority as a training and support service to help surgeries make better use of their computer systems and to improve the quality of data in primary care. PRIMIS provides training to information facilitators employed by PCTs. These facilitators then educate the GPs, nurses and other staff in effective use of their clinical computer systems and in the extraction of meaningful information. Not all PCTs in England are involved in PRIMIS and those which are, may have joined the project at different points in time. South Peterborough PCT and Huntingdon PCT have been involved with PRIMIS since the beginning, while Bedford PCT only joined it in 2001.

5.3 Identifying the Participating Surgeries

After the two rounds of workshops, the research objectives, methodology and the framework for evaluation had been clearly defined. The participating surgeries would require the information facilitators to extract data regarding the local needs, inputs, outputs and outcomes of diabetes care. Each participating surgery would also have to post a questionnaire to a random sample of their diabetic patients. All participating surgeries would receive a full report with the analyses and interpretation of the results.

Despite the interest and full financial support of the PCTs, each surgery had to be contacted in order to ascertain their willingness to participate. A letter was sent to the practice manager of each surgery, and whenever requested, a meeting with the GPs and nurses in the surgery was arranged. In South Peterborough PCT, 13 out of 15 surgeries replied positively. In Huntingdon PCT, 9 out of 25 surgeries replied positively. In Bedford PCT only one out of 25 surgeries replied positively. We were unable to ascertain why there was a very low response rate within Bedford PCT.

In April 2002, an invitation was received to present this research at a Diabetes Conference organised by Leicester Health Authority. The presentation focused on some of the drawbacks of the Government's newly published National Service Framework (NSF) for diabetes; and the DEA framework we had developed for formative evaluation in primary diabetes care. It also highlighted some of the advantages of using DEA in primary care. At this stage, we had already developed a draft version of the patient questionnaire. This questionnaire was circulated during the conference for comments (please refer to Appendix B for a copy of the patient questionnaire). After the presentation, two surgeries (one from Heartlands PCT and other from Melton and Rutland PCT) expressed an interest in taking part in this research project. A number of meetings with other health care professionals at each of these two surgeries were arranged and both surgeries agreed with the framework and methodology proposed for the project. Moreover, these two surgeries proved to have very good information systems, and became part of the sample of participating surgeries, which in total amounted to 25. Examination of the quality of the data extracted meant that two further surgeries had to be excluded, which resulted in a final sample of 23 surgeries.

5.4 Research Ethics Approval

The fact that a questionnaire had to be sent to a sample of patients meant that research ethics approval had to be sought. As this research project involved several different regions, the necessary documentation was submitted to the West Midlands Multi-centre Research Ethics Committee. The research project was given full ethics approval on 24 July 2002 (please refer to a copy of the letter in Appendix C). Following the new guidelines, all the relevant Local Research Ethics Committees were informed about the research project, and no objections were made.

5.5 Research Methodology

As we have discussed earlier, this research project aims to integrate performance assessment with evaluation research in order to develop a more comprehensive and useful performance assessment framework in primary care. In this respect, we have used a wide range of research methods to collect and analyse data, in order to gain a better understanding of the structures, practices and mechanisms behind successful primary diabetes care delivery. Additionally, we have used DEA in a practical context in order to assess its usefulness as a tool for formative evaluation.

5.5.1 The Combination of Qualitative and Quantitative Research Methods

We believe that the use of DEA in combination with other methods, such as surveys, interviews and workshops, enriches the performance assessment exercise and allows the researcher to perform a more meaningful, context-focused study. In particular, we believe that the use of qualitative methods in a quantitative performance assessment is essential for four main reasons:

(1) it gives the providers an opportunity to have an *active role* in the performance assessment exercise, by contributing to the definition of the objectives, theories, models and the interpretation of the results;

(2) it gives the researcher an opportunity to have a better understanding of the context under evaluation and the complexities involved in attempting to give it a mathematical formalisation, as well as a better understanding of the abstraction process involved in so doing;

(3) it enhances the usefulness of the results in practice and should foster organisational learning, given that both the providers and the researcher have a better understanding of the assumptions behind the results and the challenges related to its context applicability.

(4) Smith (1995) emphasised the importance of involving the providers and paying attention to non-measurable outcomes of the program, as these are essential aspects in order to minimise the unintended consequences that can result from a performance assessment exercise.

The combined use of quantitative and qualitative methods has been criticised by some authors, who believe that these two types of methods belong to opposing philosophical views of the world (paradigms), and therefore should not be used together (for example, Guba and Lincoln 1981). We believe that the debate over philosophical views does not have to mirror the debate over the most appropriate methods in a particular research context. It is our conviction that a research method *per se* does not belong to a particular philosophical paradigm. However, the researcher's philosophical positions and implicit assumptions drive the way each method is applied and how the results are interpreted. It is our conviction that all research methods have limitations and it is up to the researcher to be aware of them and to choose the most suitable methods for each research study. Cook and Reichardt (1979) recognised the importance of moving beyond the dichotomy between quantitative and qualitative methods. They urged researchers to use a combination of methods in order to best perform their specific evaluations.

We believe that the important question regarding methodology and methods is the need to have a clear conceptual framework that unifies and drives the research project. In this context, by conceptual framework we mean a set of theories about the organisations under evaluation, about what they are trying to achieve, how they are trying to achieve it, and whom they are trying to serve. We will now discuss the methodology used to develop the conceptual framework and to collect the data.

5.5.2 Methodology for Developing the Framework for Evaluation and for Data Collection

In order to develop a conceptual framework of primary diabetes care delivery, we started by performing a review of the literature and of the official documentation. This allowed the formulation of draft theories and frameworks, which provided a stepping-stone for discussion during the workshops with the PCG/Ts' Boards. These workshops identified the issues that needed to be explored and the most adequate criteria for evaluation: equity, efficiency, clinical and patient-focused effectiveness and cost-effectiveness. Once the criteria for evaluation were established, we developed the framework for evaluation and identified the data that would be necessary. It was also during these workshops that we identified what data was available and what data still needed to be collected. Two questionnaires were designed to collect the data that was unavailable: the Diabetes Services Questionnaire (DSQ) and the Practice Profile Questionnaire (PPQ). These questionnaires are discussed in greater detail in section 5.6.

The data used in the performance assessment models was extracted by the information facilitator of each PCT. After extraction, all the data was fed back to each respective surgery and meetings between the information facilitator and one of the surgeries' professionals took place in order to assess the data validity.

After the performance assessment framework was developed and validated, DEA was used to compare the sample of participating surgeries in terms of their transformation of inputs into services and services into outcomes. At the post-evaluation stage, we attempted to identify the effect of environmental variables in the performance results. Afterwards, another round of workshops with the participating PCTs and surgeries was arranged in order to validate and interpret the DEA results. This was followed by an in-depth investigation of some of the surgeries to identify how their diabetes services delivery could be improved. An assessment of the implications for primary health care delivery and primary health care policy was undertaken and is presented in chapter 10.

5.6 Questionnaire Design

We decided to use two questionnaires in order to collect data that was unavailable on the PRIMIS database. The first questionnaire was sent to a random sample of the diabetic patients registered at each surgery. The second questionnaire was sent to all 23 participating surgeries. We will now discuss each one of these questionnaires in more detail.

5.6.1 The Diabetes Service Questionnaire

When we consider that our objective was to develop a comprehensive framework for primary care evaluation, the inclusion of patient-centred outcomes is essential. Performing patient interviews was not feasible due to limitations in time. Distributing a questionnaire at the surgery could bias the results, as we needed to include the views of patients who do not regularly attend the surgery, despite being registered. For these reasons, we decided to use a postal patient survey, which allows the inclusion of a large number of subjects. The patient survey had the following objectives:

(1) to provide data in terms of the services received by the patients during the 12-month period;

(2) to provide data regarding the patients' perceived understanding of diabetes and its management;

(3) to provide data regarding patients' confidence in taking good control of their diabetes;

(4) to provide data regarding the most useful types of information, the most appropriate settings for information delivery and the 'best educators' in diabetes care;

(5) to provide data regarding the patients' evaluation of several aspects of diabetes care delivery, as well as patients' ratings regarding the importance of each aspect;

(6) to provide data regarding patients' assessment of the quality of other primary care services;

(7) to provide data regarding patients characteristics, such as age, age of diabetes diagnosis, socio-economic group, qualifications, professional position and type of diabetes treatment;

(8) to provide data regarding patients' evaluations concerning the best and the worst aspects in diabetes care delivery.

The patient questionnaire was developed in collaboration with the participating PCTs and with a number of the health care professionals at the participating surgeries. The questionnaire used by the Audit Commission (2000b) in their audit of the quality of diabetes care in England was the starting point, but several modifications in terms of content and language were made. A draft version of the questionnaire was submitted for research ethics approval in July 2002. After ethics approval was received, we piloted the questionnaire face-to-face on a sample of 20 patients from one of the participating surgeries. The patients were firstly asked if they would object taking part of a questionnaire pilot and those who did not object were interviewed in one of the consultations rooms at the surgery. This pilot process aimed at establishing if the questionnaire covered all the relevant aspects in diabetes care delivery; if the questions actually measured what they were supposed to measure; if the wording and instructions were clear to all respondents; and if the respondents felt motivated to answer it. All the patients interviewed agreed upon the clarity of the questionnaire, as well as its comprehensiveness. A few alterations regarding wording of some questions were suggested and incorporated.

Before the final version of the questionnaire was sent for Ethics Approval, it was posted to all participating PCTs and surgeries for comments. All comments suggested were duly considered. The final version of the DSQ is presented in Appendix B.

To guarantee patients' confidentiality, each surgery posted the questionnaires to the patients. Most surgeries sent out the questionnaires during October 2002. The financial burden of this exercise was met by the PCTs. A total of 3,431 questionnaires were posted and 2,201 questionnaires were received. 12 questionnaires were returned by patients who were not diabetics. This meant a total of 2,189 filled questionnaires were returned, an average response rate of 63.80%.

5.6.1.1 Diabetes Service Questionnaire Design

In order to satisfy the objectives outlined above, the DSQ was divided into four distinct sections. Section A considered the diabetic services received by the patients in the previous 12 months. The list of services covered is consistent with the British Diabetic Association (2001) 'What diabetes care to expect' in terms of the standards of diabetes care. All the registered patients should have received each of the services listed at least once over the previous 12 months. This first section also asks the patient to distinguish between the services received at the surgery and those received outside the surgery. An open question is included to allow for justifications and comments about the usefulness of the services.

Section B aimed to measure the perceived level of patients' understanding of diabetes and its management. It included a multi-item question about patients' perceived understanding of key issues in diabetes and its management, and five questions about the importance of having a good understanding of diabetes and about the most useful types and sources of diabetes information.

Section C aimed to measure patients' satisfaction with the services received at the surgery in the past 12 months. It is a multi-item question, covering several distinct aspects of quality of care, identified both from the literature and with discussions with the health care professionals.

Section D included some personal questions, such as age group, gender, ethnicity, profession, education level, age of diabetes diagnosis and type of diabetes treatment. The questionnaire ended with an open question regarding the best and worst aspects about the diabetes services provided by the surgery.

5.6.1.2 Methodology for Sample Size Calculation

The patient survey aimed at estimating measures of certain population proportions, such as: the proportion of diabetic patients who report having their blood pressure reviewed at least once during the previous 12 months; patients who report a high level of satisfaction with the diabetes services, and so on. Given the small numbers of diabetic patients registered with some of the surgeries, our initial intention was to undertake a census of all diabetic patients registered with each of the participating surgeries (5,950 patients). However, during the review process, the West Midlands Multi-centre Research Ethics Committee felt that there would be no need to survey such a large number of patients and suggested the use of a random sample. In this respect, the following criteria were used to calculate the sample size for each participating surgery:

(1) the samples taken from each participating surgery were determined using random sampling;
(2) for surgeries with a number of diabetic patients smaller or equal than 100, the whole population of diabetic patients was studied;
(3) for surgeries with a number of patients larger than 100, the formula given below was used to calculate the minimum sample size. This formula is the standard

formula used to determine the sample size of a random sample used to estimate the population proportion.

$$n \geq \frac{N z^2 p(1-p)}{d^2(N-1) + z^2 p(1-p)} \quad \text{(Source : Aiken 1997 :103)}$$

Where: N = Population size, which varies from practice to practice.

p = Proportion of the population for which the element under study is true. This value is unknown and varies for the different aspects under study. In this respect, p = 0.5 was used to calculate the sample size, given that this value gives the largest possible sample sizes.

d = Half the desired width of the confidence intervals for the proportion.

z = Reliability factor, which for this study will be 1.96, given that we aim for a confidence level of 95%.

5.6.2 The Practice Profile Questionnaire

In order to ascertain some of the characteristics of the surgeries under evaluation, a questionnaire was sent to each participating surgery. A copy of the practice profile questionnaire (PPQ) is presented in Appendix D. The PPQ was modified from a questionnaire developed by Williams et al. (2000) to assess the structural provision of diabetes services in the UK. The main objectives of this questionnaire were:

(1) to provide information about the size of each surgery, its location and its characteristics;

(2) to provide information about the number of professionals it employs, including the number of professionals that regularly visit the surgery;

(3) to provide information about the frequency of diabetes related consultations, as

well as their duration;

(4) to provide information about the diabetes screening policy of each surgery;

(5) to provide information about the educational materials distributed to the diabetic patients;

(6) to provide information about the training of the health care professionals providing diabetes services at the surgery.

A PPQ was received from each surgery, despite the fact that some surgeries were unable to complete some of the sections of the questionnaire.

5.7 Conclusion

In this chapter we have discussed the methodology used to develop the framework for evaluation in primary diabetes care and the methodology used to collect the data. When we consider that this study aims to contribute to formative evaluation in primary care, an emphasis has been given to the participation of the providers and to the development of a framework that portrays alternative performance criteria. As a result, the data collection has been driven by theories in terms of primary care delivery and by the values behind decision-making in primary care.

In order to obtain meaningful results that can be useful both for the health care professionals and the patients, we have complemented the data collection with the use of two questionnaires. The first was sent to a sample of patients to get the patients' views regarding the delivery of primary diabetes care. The second was sent to all the surgeries to get information to characterise each one of the participating surgeries.

In the next chapter we will discuss the performance assessment framework for diabetes service delivery.

6 The Framework for Formative Evaluation in Primary Diabetes Care

6.1 Introduction

In this chapter we will provide the background for the use of DEA for formative evaluation of primary diabetes care delivery. In Section 6.2 we justify the need to perform a formative evaluation in primary diabetes care. In section 6.3 we discuss the drawbacks of the National Service Framework (NSF) for diabetes developed by the Government. A definition and characterisation of this chronic condition is provided in section 6.4. In Section 6.5 we review the previous studies that have compared the performance of primary care providers in terms of diabetes care delivery. In section 6.6 we discuss the framework developed for evaluation of primary diabetes care delivery, which entails a discussion of the needs, objectives, inputs, outputs and outcomes that characterise primary diabetes care delivery in England.

6.2 Why do we need a Formative Evaluation of Primary Diabetes Care Providers?

> "This is a crucial time for diabetes services. The numbers of people with diabetes are increasing, new evidence is emerging on the effectiveness of more intensive treatment, patient expectations are rising and services are being stretched to the absolute limit. The

question is, how can services cope with this rising demand?"

(The Audit Commission 2000b: 8)

There are wide variations in the standard of the diabetes care provided across the country (The Audit Commission 2000b). However, to our knowledge, no published study has compared a sample of primary health care providers in terms of equity of access to services, efficiency and effectiveness; this is the objective of this case study. In order to identify the structures and mechanisms associated with *best practice*, we performed a comparative evaluation of the performance of 23 GP surgeries in terms of their delivery of primary diabetes care. Primary diabetes care refers to the comprehensive range of health services provided to patients in order to diagnose diabetes, give support to the diagnosed patients and minimise the consequences of diabetes. It excludes services that require patient hospitalisation.

With so many specialities to benchmark, it would be prudent to ask why we have chosen diabetes. The answer is simple, the rising prevalence and health care costs associated with this condition. In the UK, current diabetes prevalence amounts to over 3% of the population, with another 3% of the population estimated to have undiagnosed diabetes (The Audit Commission 2000b). The total costs of diabetes care to the NHS are estimated to exceed 2 billion pounds a year (The Audit Commission 2000b), and are expected to increase significantly over the next decades.

Although diabetes care is currently at the top of Governments' agenda, and is receiving a great deal of attention in the literature, to our knowledge, no published study has concentrated on the performance of primary diabetes delivery systems using a comprehensive framework in a formative mode. Therefore, the identification of the structures and mechanisms associated with an equitable, efficient and effective delivery provides essential information for both health care professionals and policy makers. Furthermore, we believe that this case study is very timely, because a new GP contract has been approved stipulating increased pay for the GP surgeries if certain performance targets are achieved (British Medical Association 2003). Diabetes care is one of the specialities subject to performance related pay. Each GP surgery will have to decide whether it is cost-effective to deliver diabetes care, and if so, under what model. Structure, process and outcome targets have been defined, with greater weighting being given to outcomes achievement (British Medical Association 2003). Given the Government's emphasis on clinical outcomes achievement, we believe that it is important to investigate what factors may be associated with clinical and cost effectiveness, as well as discuss the potential unintended consequences that may result from a target driven system of diabetes care delivery.

6.3 The Drawbacks of the National Service Framework for Diabetes

> "The NSF aims to make best practice the norm: ... community-based diabetes clinics to bring together at one location and at one time specialist expertise and services, including foot care and eye screening. ... Alongside the real progress that has taken place in some areas and for some places, there remain significant variations in the quality of care. Excellent diabetes services in one place can exist cheek-by-jowl with diabetes care elsewhere that is inadequate and unimaginative. Recognition of these challenges led the Government to initiate the development of this *National Service Framework for Diabetes*." (The Department of Health 2001a: 2, 3)

The NSF for diabetes sets out the standards that must be achieved in diabetes care delivery in the NHS. The strategy for delivery of these standards has also been outlined (The Department of Health 2003). The table presented in Appendix E provides the list of standards outlined in the NSF for diabetes. To monitor the achievement of these standards a large set of PIs have been developed and targets have been set for these PIs. Each provider will have to implement the necessary changes in order to achieve these targets.

However, in our view, the NSF misses a crucial element in diabetes care: costs. The NSF does not include any indicator linking the resources used in diabetes care with the processes and the outcomes achieved. Without this crucial link, strategies to improve efficiency and cost effectiveness in diabetes care cannot be identified. To assess providers on the achievement of outcome targets for a particular speciality, like diabetes, without taking into account the delivery costs is problematic. It can lead to problems of affordability, as well as, lead to resource shifts away from specialities for which no targets have been imposed. In a context where resource rationing is unavoidable, it is of paramount importance to identify and investigate the systems of care delivery, which are able to produce more/ better outcomes, with the same (or lower) costs. It is this essential link between needs, resources used, processes of care, services delivered and outcomes achieved that we attempt to restore by using DEA. We discuss the results of our empirical study in chapter 7. We now provide a definition of diabetes.

6.4 *Diabetes Mellitus* – Definition and Characterisation

Diabetes Mellitus is a progressive chronic condition in which the body is unable to effectively control the amount of sugar in the blood, allowing it to rise to abnormal levels. It is commonly known as diabetes. This condition is currently incurable, and shows rising prevalence in the UK. The number of diagnosed diabetics is estimated to rise significantly during the next decades, partly due to better ways of screening for the condition, but mainly due to a widespread change in lifestyle towards a high fat, high-energy diet, combined with low physical activity.

There are two types of diabetes:

(1) Type-1 (formerly known as Insulin-Dependent-Diabetes), is caused by a deficiency of the pancreas to produce insulin. For type-1 diabetics the insulin producing cells (β-cells) have been destroyed by the immune system. In this respect, type-1 patients depend upon the daily injection of insulin to survive. The symptoms of this form of the condition tend to develop very rapidly and its diagnosis is more common in children and young adults.

(2) Type-2 (formerly known as Non-Insulin-Dependent-Diabetes) is caused by the pancreas inability to produce enough insulin, associated in most cases with an inability of the cells to use the small quantities of insulin produced. In this case, the patient is treated with any combination of diet, medication and insulin. This type of diabetes is the most common form of the disease, representing around 85% of the cases in the UK. Its symptoms tend to develop more slowly, and the diagnosis usually occurs in patients over 40 years of age. However, type-2 diabetes has been increasingly reported in children and adolescents (American Diabetes Association 2000), coinciding with the increasing prevalence of obesity.

Although desirable, it was not possible to extract separate data based on which condition afflicts the patient, because the coding of the two types of diabetes is still not standardised between surgeries. At a time when this type of data extraction is possible, we recommend that separate analyses should be made for type-1 patients and type-2 patients, considering that they require slightly different types of treatment, and the progression of the condition is also slightly different. Based on the data from the patient questionnaire, we have estimated the proportion of patients with type-1 diabetes. The criterion used to classify type-1 patients was: age of diagnosis < 41 years old and treatment with insulin. The estimated proportion of patients with type-1 diabetes for each surgery is later used as an environmental variable to explain the efficiency and effectiveness results obtained.

Both types of diabetes can lead to early death, poorer quality of life and the onset of serious health complications, such as: cardiovascular diseases, foot ulceration, kidney failure, eye diseases and blindness. Research suggests that an optimum metabolic control (blood glucose, blood pressure and cholesterol) can effectively prevent or delay these associated complications, both in patients with type-1 diabetes and in patients with type-2 diabetes (Wake et al. 2000). The outcome measures used in this study reflect this research evidence.

6.5 Previous Studies that have Compared Primary Care Providers in Terms of Diabetes Services Delivery

Most research in diabetes care has focused on the effects of different types of medication, on the development of complications and the associated health care costs. The standard method used in these studies is a randomised control trial, where a random sample of patients is subjected to a new type of medication and a control sample is subjected to a conventional type of medication. The costs and outcomes of the patients from both samples are compared in order to infer about the cost-effectiveness of the new therapy.

To date, four major studies have been published on this subject. The Diabetes Control and Complication Trial Study Group (DCCT) (1993) compared the costs and outcomes of type-1 patients under standard treatment and patients under insulin intensive treatment. This study found that intensive insulin therapy contributed to a reduction in the development and progression of microvascular complications in patients with type-1 diabetes. The comparison of the costs involved in the two types of treatment suggested that intensive treatment is cost-effective in the long run.

The Stockholm Study (Reichard 1992) reported similar findings, although the cost figures were significatively lower. These findings have now been extended to patients with type-2 diabetes. Two major studies have been undertaken in the UK and Japan: The UK Prospective Diabetes Study Group (UKPDS) (1998a) and the Kumamoto study (Wake et al. 2000). These two studies found that better metabolic control by intensive insulin therapy can significantly reduce and delay diabetes related complications. Despite the increase in treatment costs due to the use of insulin, intensive therapy was found to be cost-effective in the long term due to a reduction in the costs associated with the treatment of complications. Smaller scale studies have investigated the effects of different types of medication on the cost-effectiveness of treatment. A full review of these studies is outside the scope of this thesis. The objective of this case study is to investigate other types of factors that can explain the efficiency and cost-effectiveness of primary diabetes care delivery. In particular, we are interested in investigating provider-related factors.

There are few studies that have investigated diabetes care performance at the provider level. Most of the studies that have performed analyses at the provider level suffer from some limitations. In particular, to our knowledge, no published study has investigated the relationships between equity, efficiency and effectiveness of primary diabetes services providers. The research emphasis has been on the quality of care under different delivery systems without considering the delivery costs.

The Diabetes Integrated Care Team (1994) undertook a study involving the random allocation of 274 diabetic patients into conventional hospital care and integrated care[3]. The aim of the study was to investigate the effect of type of care on certain outcome measures, such as metabolic control, knowledge of diabetes, satisfaction with treatment, psychosocial status and costs to the patients and to the NHS. The study concluded that after two years, the outcomes of patients allocated to integrated care were as good as those allocated to conventional hospital care. The costs to the NHS were found to be similar, but integrated care imposed lower costs to the patients, particularly for patients living far from the hospital. This early study pointed towards the provision of diabetes care in primary care, given that it appears to be more convenient to patients, while maintaining cost-effectiveness. In recent years, there has been a shift from secondary to primary care in terms of the management of most chronic conditions, including diabetes (Goyder et al. 1998).

Griffin (1998) reviewed all published randomised control trials that evaluated outpatient hospital clinics versus general practice care for diabetes. He concluded that, if a computerised prompt system exists, general practice could be as effective as hospital outpatient clinics in the delivery of diabetes care. From another aspect, he concluded

---

[3] Patients allocated to integrated care were seen annually at the hospital, and periodically at the GP surgery. Patients allocated to conventional hospital care were seen periodically at the hospital.

that when a suitable organisation system of diabetes care is not in place, general practice care can lead to poorer services and outcomes.

More recently, Ovhed et al. (2000) investigated the importance of the role of the diabetes nurse in primary diabetes care in Sweden. The study compared two Primary Health Care Centres (PHCC) operating different models of diabetes care. PHCC A had implemented formal guidelines for diabetes care, establishing an independent role for the diabetes nurses. In this centre, the prerequisite of diabetes care was that patients should be seen three times a year with the nurse and once a year with the doctor. PHCC B had no formal guidelines for diabetes care and the doctor took the main role in diabetes care, seeing all patients twice a year. Diabetes nurses acted as assistants and the patients were only sporadically referred to a nurse when the doctor felt that it was necessary. The study concluded that PHCC A, where the nurses had an independent and more active role, presented better metabolic control. The authors recognised that this aspect needs to be further investigated particularly by comparing a larger number of primary care providers.

Glasgow and Strycker (2000) undertook a comparison of 47 physicians from two different health care systems in the United States. They compared the physicians across 11 performance measures, which assessed the percentage of patients that had the essential laboratory tests in the previous year, and the percentage of patients that had received patient education of diabetes management in the previous year. A great deal of variability in the performance measures across providers was found. In particular, patient education was found to be quite unsatisfactory in most providers. In a second stage, multiple regression analyses were used to identify the factors associated with the performance variability. The results from the 11 performance measures for each physician were regressed on a set of patient and provider characteristics. They concluded that several variables were associated with better performance, such as the number of patients treated and the use of guidelines. One interesting finding was that the level of community support for the patients (support from family, friends, work, and other community resources) was found to be the only significant predictor of better performance across both health care systems.

Despite the interesting findings of this study, it suffers from some limitations. Firstly, it would have been interesting to perform analyses at the practice level. Secondly, the authors used a simple average to aggregate all the performance measures. This is equivalent to giving equal weights to all the measures across all providers. An approach allowing some weight flexibility would be fairer, given that it would allow each provider to emphasise its specific strengths. Finally, and most importantly, no conclusions can be taken regarding the relative efficiency and cost-effectiveness of care because the authors did not link the outcome measures with any measure of cost.

Another study that fails to include an assessment of the costs involved in the delivery of diabetes care was that by Khunti et al. (2001). They compared a sample of 169 practices in England in terms of some process and outcome measures in diabetes care. A wide variation across practices was found in both process and outcome measures. Additionally, most practices showed very poor levels in terms of outcomes achievement. The authors performed multiple regression analysis in order to identify the factors associated with the performance variations. Four factors explained a small part of the variability:

(1) the size of the practice (smaller practices performed better);

(2) fundholding status (fundholding practices performed better);

(3) the existence of a recall system (practices operating a recall system performed better, but only in terms of the process measures);

(4) the proportion of patients under hospital diabetes care (practices with a smaller proportion of patients under hospital care performed better).

Given the general poor quality of care identified, the authors suggested that the current payment for chronic disease management might be insufficient to guarantee high quality of diabetes care.

The isolated study of costs and outcomes of diabetes care has been one of the major limitations of most studies in this research area. The majority of studies of diabetes care performance have either focused on outcomes, whilst neglecting the costs, or focused on costs, without linking it to the outcomes achieved. We will now discuss the development of our framework for formative evaluation in primary diabetes care.

6.6 Developing a New Framework for Formative Evaluation in Primary Diabetes Care

We now discuss the framework used for formative evaluation in primary diabetes care. This framework follows the generic conceptual framework presented in chapter 4. In this respect, we have used the following conceptual representation:

Needs ⟶ Objectives ⟶ Resources ⟶ Service delivery ⟶ Outcomes

By comparing the level of services delivered across surgeries to the essential services that the patients need, we get a measure of relative equity across surgeries. By comparing the quantity of appropriate services delivered to the amount of inputs used we get a measure of relative efficiency across surgeries. By comparing the amount of services delivered to the level of outcomes achieved we get a measure of relative effectiveness across surgeries. Finally, by comparing the level of outcomes achieved to the amount of inputs used we get a measure of relative cost effectiveness.

The period of analyses is 1 October 2001 to 30 September 2002. This was the most recent period for which data was available at the time of extraction. The final framework used for evaluation is presented in Figure 6.1. The final measures used in the models, resulted from a consideration of the following aspects:

(1) importance for the formative evaluation exercise, based on the literature and workshops with the health care professionals;

(2) fair attribution of the performance results. For example, to monitor the incidence of diabetes related complications would not be appropriate, when we consider that most of these complications are related to the diabetes care provided many years prior to the onset of the complications. The number of patients with optimal metabolic values is a better indicator of the quality of diabetes care regarding the prevention of complications, because it captures the patient's current risk of developing complications and is influenced by recent medical care or its absence;

(3) availability of reliable and consistent data across the surgeries, or possibility of collecting reliable data using the patient questionnaire;

We will now discuss in more detail each of the elements involved in this formative evaluation, in order to justify the measures used.

6.6.1 Local Needs of Primary Diabetes Care

In Chapter 4 we have discussed the concept of need. As suggested by Williams (1974) and Culyer (1976), we have equated need for health care with the patients' capacity to benefit from the consumption of health care. The fact that we have narrowed our focus from primary care as a whole to primary diabetes care facilitates the assessment of need. For our empirical analysis, we have defined two groups of patients with need for primary diabetes care:

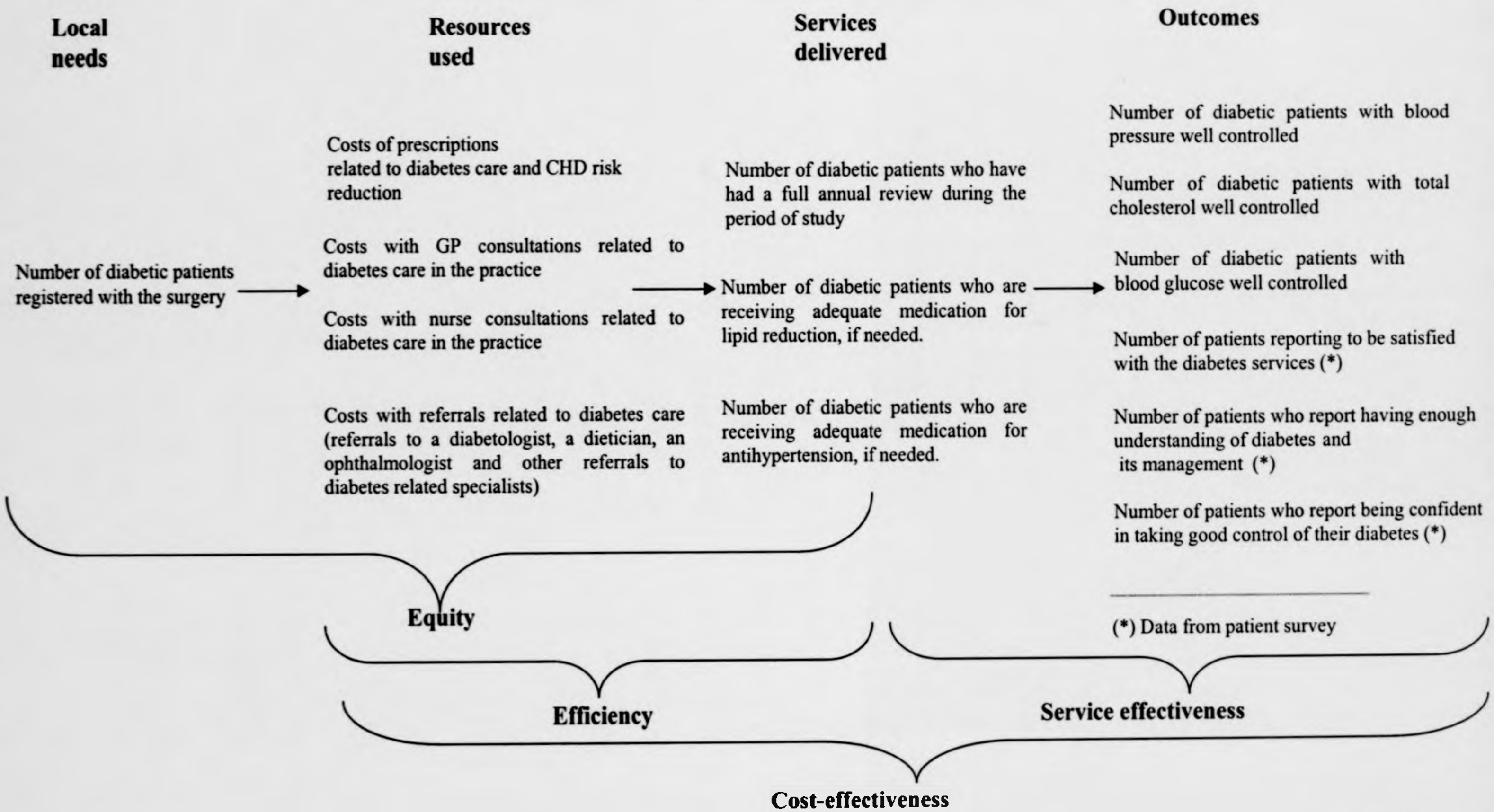
(1) Patients most at risk of developing diabetes, who can benefit from a screening exercise;

(2) Patients already diagnosed with diabetes, who can benefit from a comprehensive annual review and appropriate diabetes medication;

Most of the participating surgeries do not follow a regular programme of diabetes screening. As a consequence, we have focused solely on the delivery of diabetes care to patients already diagnosed with diabetes. The number of diabetic patients registered with each surgery is therefore our proxy measure for local need. We recognise that this is a crude measure of need given that it does not discriminate between patients with different levels of severity, nor does it distinguish between type-1 and type-2 diabetics. Nevertheless, there is a range of essential services that all diabetics should receive at least once a year (British Diabetic Association 2001). Therefore, we believe that it is useful to investigate if the proportion of patients who receive these essential services varies significantly between surgeries.

Figure 6.1: Framework for performance assessment in primary diabetes care

Level of analysis – The Surgery



6.6.2 Objectives of Primary Diabetes Care Delivery

The main objectives identified in the literature (St Vincent UK Working Group 1996 and Home et al. 1999) and corroborated in the workshops with the PCTs were:

(1) To reduce the incidence of diabetes, through preventive action.

It is not thought to be possible to prevent the onset of type-1 diabetes. However, given that obesity triggers the onset of type-2 diabetes, it is thought that the use of measures aimed at reducing obesity and increasing physical activity could contribute towards a reduction in the incidence of type-2 diabetes. It would also be an effective measure of heart disease prevention.

Despite the importance of diabetes prevention, this study will not focus on this type of service, because at the time of this study, most surgeries were not actively involved in diabetes preventive care. Once this becomes an integral part of the regular activities of GP surgeries, we recommend that the costs, services and outcomes associated with preventive care are also included in a new performance study. An analysis of the observed trade-offs between preventive care and disease management care could provide interesting insights into the optimal balance between these two areas.

(2) To detect diabetes as early as possible, through community education on symptom recognition and through regular screening.

Type-2 diabetes screening is recommended by the World Heath Organisation (1994). In the UK, Wareham and Griffin (2001) review the evidence and suggest that although universal screening is not recommended, targeted screening appears to be cost-effective. It is possible to provide patients with urine test sticks to screen at home for diabetes. This test, if accompanied by clear instructions, could be a cheap and effective method to screen most at risk patients. However, any type of systematic screening would obviously identify a large amount of unmet need, for which the primary care system may not be prepared.

Despite this important short-term problem of insufficient capacity, an early diagnosis is essential both in terms of patients' expected quality of life and long-term cost savings to the NHS. It is believed that half of the patients diagnosed with type-2 diabetes already have complications at diagnosis (UKPDS 1990).

(3) To reduce the risk of complications of diabetes, through regular monitoring and by using adequate medication.

In order to reduce the risk of complications in diabetes, the blood glucose concentration has to be tightly controlled. Moreover, given the fact that diabetes tends to lead to the development of cardiovascular diseases, it is important to control the cardiovascular risk factors, such as smoking, obesity, blood pressure (BP) and cholesterol levels.

(4) Effective patient education in diabetes and its management, through regular consultations with health care professionals, and the provision of adequate education materials.

It is essential to recognise that patient education and patient compliance are essential in achieving good metabolic control. It is also very important to recognise that patients with chronic conditions often have a wealth of knowledge and understanding about the optimal day-to-day management of their condition. Enhancing this practical wisdom and supplementing it with clinical knowledge is supposed to make the patient a key element in the achievement of good outcomes of care (The Department of Health 2001b). Obviously, this discourse regarding the benefits of patient education and patient empowerment assumes that the patient both wants and has the capability to take responsibility for disease management.

(5) To reduce the impact of diabetes and its complications once developed, through appropriate patient support and effective treatment.

Diabetes is the leading cause of kidney failure, blindness and one of the biggest causes of lower limb amputation. Diabetes pre-disposes to cardiovascular diseases. A significant proportion of diabetic patients require psychological support in order to overcome depression and anxiety (Lloyd, Dyer and Barnett 2000). It is essential that a regular and effective programme of care be in place to prevent these health problems, or to minimise its consequences.

6.6.3 Inputs – Resources used in the Delivery of Primary Diabetes Care

In this study we have used a combination of methods in order to arrive at estimates for the resources used by the GP surgeries in their delivery of diabetes care. From one side, we collected data from the GP surgeries in terms of the resources used at the surgery in diabetes care. These include the Whole Time Equivalent (WTE) GP, the WTE nurse and the number of medication items prescribed. A questionnaire was sent to a large random sample of patients with questions about the use of resources outside the surgery. We used the answers to the patient questionnaire to estimate the number of consultations with a diabetologist, an ophthalmologist and a dietician.

The input data extraction from the surgeries' registers proved to be one of the most difficult aspects of this case study, because the surgeries do not operate a speciality-based-costing. In order to identify the costs attributable to primary diabetes care, new extraction queries had to be written and the data for most inputs had to be extracted for the first time. Additionally, a number of assumptions had to be made regarding the number of consultations with health care professionals. The difficulty arises because most patient consultations have more than one topic of discussion, and that health care professionals do not yet code consultations detailing the main topic of discussion. Therefore, we had to assume for all the surgeries that any diabetic patient consultation was an opportunity for ongoing diabetes monitoring, patient education and support. This was considered to be the most appropriate assumption after discussions with the health care professionals. It was their opinion that most consultations cover several topics and that diabetes is often a topic of discussion in any consultation with a diabetic.

The data regarding the costs with diabetic medication was obtained from the Prescribing Analysis and Cost (PACT) system published electronically by the Prescribing Price Authority (PPA). We now discuss in more detail each one of the inputs involved in primary diabetes care delivery.

6.6.3.1 Diabetes Screening

The early detection of diabetes is an important objective of primary diabetes care. As previously discussed, it is estimated that around half of the type-2 patients already have complications at the stage of diagnosis. If an earlier diagnosis can be made, significant savings can result. Therefore, regular screening is one of the policies recommended in the diabetes NSF. However, at the time of study, regular screening was not yet policy. We have decided to exclude this information from our analyses. Once active screening becomes common practice, its inclusion in a performance assessment exercise is recommended.

6.6.3.2 Medication Related to Diabetes Treatment

In order to measure technical and allocative efficiency, we have collected data in terms of quantities and unit costs for the different medications related to diabetes treatment.

The data for the diabetic medication is provided directly from the PACT system because these items relate to drugs and materials only prescribed to diabetics. The costs in these categories can be attributed in full to the registered diabetics. However, the data for the cardiovascular risk reduction medication is not as straightforward to extract. Both lipid-lowering drugs (for example, statins) and Angiotensin-converting Enzyme Inhibitors (ACEI)[4] are items that can also be prescribed to non-diabetics. In this respect, given that we intend to identify the prescribing costs solely attributable to diabetes, a different method of data extraction had to be used. Firstly, we extracted the number of diabetic patients on Lipid-lowering drugs and the number of diabetics on ACEI. Each 28 days of treatment for each patient counts as one item of medication. We then used the PPA (2002) average cost estimates for each one of these items[5].

6.6.3.3 Laboratory Tests

A large number of regular blood, urine and other tests have to be done with diabetic patients. This represents a significant proportion of the resources used in primary diabetes care. However, despite our efforts, no reliable estimates could be obtained for

[4] Angiotensin-converting enzyme inhibitors are medicines that block the conversion of the chemical 'angiotensin I' to a substance that increases salt and water retention in the body.

[5] These average costs represent the costs of treating a patient for 28 days on the adult recommended dose of the most frequently prescribed drug for the purpose.

the number of tests carried out by each surgery, given that the surgeries are not charged per volume of laboratory tests. A block contract exists between the PCTs and the laboratories and the payment is independent of the number of tests requested. Therefore, this aspect does not constitute a performance incentive. Unfortunately, duplication of tests and procedures can be one of the causes of technical inefficiency. We recommend that this component is included as soon as reliable information can be extracted.

6.6.3.4 Consultations with the GP, Diabetologist and Nurse

The extraction of data regarding the number of consultations requires the surgeries to keep an accurate record of all consultations, by patient number. Therefore, we had to restrict our sample of surgeries to those that operated a fully computerised recording system. We extracted data regarding all the consultations with diabetics that took place during the period of study. These consultations were then disaggregated by place of consultation (surgery consultation, home visit, telephone consultation) and by consultant (GP and nurse).

The number of consultations with a diabetologist could not be extracted directly from the surgeries' records. Furthermore, some of the surgeries were not able to provide an estimate of the proportion of patients under diabetes hospital care and under diabetes shared care[6]. Therefore, we decided to use the data collected from the patient survey to estimate the average number of consultations with a consultant at the hospital during the period of study[7].

---

[6] Patients under shared diabetes care are regularly consulted both at the GP surgery and the hospital.

[7] We used the average number of times that patients reported to have had a consultation with a doctor outside the surgery to estimate this value.

In order to obtain information regarding the average duration of the consultations with a GP and a nurse, a questionnaire was sent to each surgery (Practice Profile Questionnaire – Appendix D). We intended to use this information to calculate the WTE GP and WTE nurse. However, not all surgeries completed this section of the questionnaire. Therefore, we used 'a national average duration of consultations' based on the figures published by Netten, Rees and Harrison (2001). The unit costs of one hour of patient contact with a nurse and a GP are also based on the estimates obtained by Netten, Rees and Harrison (2001). Table 6.1 presents the average duration of each type of consultation and the costs per WTE hour, as used in this study.

6.6.3.5 Consultations with other Diabetes Related Specialists

We wanted to include data regarding the number of times each diabetic patient had been seen by a diabetes-related specialist. However, due to a lack of coordination in the data register systems, reliable figures could not be obtained. We have used the responses from the patient questionnaire (Appendix B, section A) to estimate the average number of times that patients had a consultation with an ophthalmologist and a dietician.

Unfortunately, we could not estimate the number of times patients had a consultation with a chiropodist because some patients mention that they regularly consulted a private chiropodist (this is a common feature in terms of foot care, given the long waiting times under the NHS). Therefore, we have decided to exclude the costs with the chiropodist from the models. We recommend that this input is included in future studies, once reliable data can be extracted. Table 6.1 presents the unit costs used for a consultation with an ophthalmologist and a dietician.

Table 6.1: Average duration of consultations and costs per WTE hour

	Average duration of consultations in minutes[8]			Cost per WTE hour[9] (2000/2001 prices)
	Clinic consultation	Home visit (Includes travelling time and travelling expenses)	Telephone consultation	
GP	12.56	29.30	10.47	£86
Practice Nurse	18.26	26.09	10.47	£23
Diabetologist	-	-	-	£ 74 per outpatient appointment
Ophthalmologist	-	-		£ 74 per outpatient appointment
Dietician	30	-	-	£27

6.6.3.6 Smoking Cessation Therapy

Smoking is a cardiovascular risk factor to the general population. A number of research studies have suggested that the combination of smoking and diabetes leads to the premature development of a range of health complications, because smoking increases insulin resistance and interferes negatively with insulin action. Haire-Joshu, Glasgow and Tibbs (1999) review this literature and suggest that effective smoking cessation therapy should be a routine component of primary diabetes care. In the UK, smoking cessation is a major priority in primary diabetes care delivery and smoking cessation

[8] The average duration of the consultations for GPs was obtained by dividing the cost of that type of consultation by the cost of one minute of GP contact with patients. The average duration of a nurse home visit and a nurse clinic consultation were calculated in the same way. The average duration of a nurse telephone consultation was assumed to be the same as for a GP.

[9] The unit cost for the GP consultations includes overhead costs. The unit cost of an outpatient visit refers to the cost of a generic visit, which is an average of the costs of a wide rage of specialities (Netten, Rees and Harrison 2001: 83). We used this estimate because specific estimates for diabetologist and ophthalmologist consultations could not be obtained.

services are currently being developed in most GP surgeries. However, at the time of data extraction, reliable data regarding the costs and outcomes of these programmes could not be obtained. Therefore, we have decided to exclude this item from the models. We do recommend the inclusion of smoking cessation measures once reliable data becomes available.

6.6.4 Outputs – Services Delivered in Primary Diabetes Care

The common output measure in studies of efficiency in health care provision is the age-standardised number of registered patients. We argue that this is a poor output measure, when we consider that a registered patient who has not received any service during the year of study should not be counted as an output. Therefore, we suggest the use of an activity measure. In this study, the outputs used reflect 'services delivered in the year of study'. We now discuss in detail each of the outputs used in this study.

6.6.4.1 Diabetes Annual Review

Measure: Proportion of patients who attended a diabetes annual review (or initial review) during the period of study.

A full annual review is the most basic service that surgeries should deliver to their diabetic patients. This review should include a range of tests and diagnostic procedures, a review of the patient's diet and/or medication, a discussion with the patient regarding his/her wellbeing and a discussion of the results of the tests. It is essential that all the information of the annual review be electronically recorded, as this is the basis for an automated system of patient call and re-call. Automatic letters can be produced and sent to the patients before the 12-month period is completed, and the system can be designed to give warning messages whenever a patient has not been seen for more than a pre-defined period. The proportion of registered diabetic patients who are annually reviewed either at the surgery or at the hospital should therefore be very high[10]. The records of the patients who are not annually reviewed should be closely inspected, and these patients should either be invited for a review, or removed from the records if they are no longer registered with the surgery or if they are found not to be diabetics. However, most surgeries fall very short of this performance measure (The Audit Commission 2000b), either because they are not reviewing their patients often enough, or because they are not recording it.

There is a standard register code to be entered into the patient's medical records each time he/she is fully reviewed. However, some health care professionals either use their own codes or do not insert this information electronically. In this study, an effort was made to extract data including all the codes in use at each surgery. We have assumed for all the surgeries that a recorded annual review includes all the tests that are recommended as well as a review of medication and education. There was no way of verifying this from the surgeries records. The proportion of registered diabetic patients who have had a full annual review is, therefore, the first, and most important, output measure in primary diabetes care delivery. The number of patients who have had an annual review includes the newly diagnosed patients[11] who have had a full initial diabetes review.

---

[10] The new GP contract (British Medical Association 2003) establishes a target of 90% for this measure.

[11] Newly diagnosed patients are those patients who were diagnosed as diabetics during the period of study, that is, between October 2001 and September 2002.

6.6.4.2 Cardiovascular Risk Reduction

Measure 1: Proportion of patients with cardiovascular diseases or at risk of developing cardiovascular diseases, who are on lipid lowering medication.

Measure 2: Proportion of patients aged 40 years old or more who are on anti-hypertensive medication.

In order to account for the services related to cardiovascular risk reduction, we have decided to include two outputs regarding the proportion of high-risk diabetic patients who are under adequate medication.

The first of these output measures represents the proportion of patients with Ischaemic Heart Disease (IHD), or at risk of developing IHD (that is, diabetic patients aged between 35-70 years old, with total cholesterol > 5 mmol/l[12]), who are on regular treatment with lipid lowering medication. The most common type of medication for lipid lowering is statins. The prescription of statins to diabetic patients has been showed to lead to significant reductions in coronary events, revascularisation and stroke (Heart Protection Study Collaborative Group 2003).

The last output measure represents the proportion of diabetic patients aged over 40 years old who are on anti-hypertensive medication. The most common type of medication used is ACEI. Regular treatment with ACEI is recommended in all diabetics over 40 years old, independently of their BP value, because ACEI have been shown to have other major health benefits in diabetics (cardiovascular and renal protection), independently of BP reduction (The Heart Outcome Prevention Evaluation Study Investigators 2000).

---

[12] Mmol/l is the standard unit of measurement of concentration of cholesterol in the blood.

6.6.5 Outcomes: Impact of Primary Diabetes Care

The most important outcomes of primary diabetes care relate to the achievement of near-normal metabolic values, the improvement of patient understanding of diabetes and the improvement of patients well being. The specific outcome measures chosen for this study, as well as the specific threshold levels for the clinical measures resulted from the literature review and the discussions with the health care professionals. Measures of outcome achievement were chosen instead of outcome improvement because these were believed to better reflect how close each surgery is to the achievement of its objectives. Furthermore, the new GP contract (British Medical Association 2003) also uses measures of outcome achievement for target setting. We provide a brief justification for the specific measures used, as well as the specific threshold levels chosen.

6.6.5.1 Blood Glucose Level

Measure: Proportion of patients with the latest Glycated Haemoglobin (HbA1c) ≤ 7% (it requires the patients to have at least one HbA1c measurement during the 12-month period of study).

Glycated Haemoglobin (HbA1c) measures the average blood glucose level of the previous 120 days, with a greater weight given to the 30 days preceding the measurement. This is a standardised measure of 'long-term' blood glucose, used by researchers, health professionals and patients to identify targets of diabetes control.

The generic aim is to reduce the levels of blood glucose to as near normal levels as possible. The normal level of HbA1c in non-diabetics is between 5.5% and 6.5%. The results of the UK Prospective Diabetes Study Group (UKPDS) (1998a) suggest that diabetes related complications and diabetes related death can be prevented by lowering blood glucose levels in patients with type-2 diabetes.

Using the UKPDS (1998a) recommendation as a template, it was agreed that a 7% measure should be used as the cut-off point, despite the use of a 7.5% cut-off point in the NSF for diabetes. An HbA1c of less than 7% is considered a good outcome, whilst an HbA1c between 7% and 7.5% is only considered as an acceptable outcome.

6.6.5.2 Blood Pressure Level

Measure: Proportion of patients with the most recent reading of Blood Pressure (BP) ≤140/80 mm Hg (it requires the patients to have at least one BP reading during the 12-month period of study).

The inclusion of this outcome results from strong research evidence supporting a tight control of BP in diabetic patients. There are two major studies supporting a tight control of BP in diabetic patients. The UKPDS hypertension study (1998b) involved a randomisation of a sample of diabetics with hypertension into conventional treatment and intensive treatment. A significant reduction in diabetes related complications and death was observed in patients subject to intensive BP control. The second study was the Hypertension Outcomes Trial (Hansson et al. 1998), which involved the randomisation of a large number of hypertensive patients into three groups (aiming for diastolic BP under 90, under 85 and under 80). This study showed a 50% reduction of cardiovascular morbidity and mortality in the diabetic patients achieving diastolic BP under 80 mm Hg.

6.6.5.3 Blood Cholesterol Level

Measure: Proportion of diabetic patients with the most recent measurement of total cholesterol ≤ 5 mmol/l (it requires the patients to have at least one total cholesterol measurement during the 12-month period of study).

Total cholesterol concentration is a risk factor for the development of CHD within the general population. The results of the UKPDS (1998a) suggest that an increased concentration of total cholesterol is a major risk factor in the development of CHD in patients with diabetes type-2. As secondary prevention, therapy is recommended to reduce the total cholesterol level in all patients with history of CHD. It is also recommended as primary prevention in diabetic patients who are at risk of developing cardiovascular complications. The threshold of total cholesterol ≤ 5 mmol/l reflects the research evidence and the views of the clinicians as to what level is desirable.

6.6.5.4 Patient Understanding of the Condition and its Management

Measure: Average proportion of patients who reported to have a good understanding in terms of diabetes and its management[13].

Patients' motivation and ability to effectively manage the condition is crucial to achieve optimum metabolic control and to prevent the development of complications (Home et al. 1999). Health care professionals play a crucial role in terms of patient education and empowerment. Therefore, we believe that the level of patient understanding of diabetes and its management is an important outcome of primary

---

[13] This measure was estimated by calculating the average of the proportions of patients who answered 'I understand enough' to items a-i included in Section B of the questionnaire (Appendix B).

diabetes care. In this study we have used a multi-item question to assess this outcome. This question assesses the patients' perceived understanding of the essential aspects of diabetes and its management. These are essential aspects related to diabetes, its management and its effects on a patient's health. Any patient with diabetes should have a reasonable understanding of all these aspects. For the list of questions used, please refer to the patient questionnaire (Appendix B, section B).

6.6.5.5 Patient Perceived Quality of Diabetes Care

Measure: Average estimated proportion of patients who answered 'Always' or 'Most Times' regarding the quality of diabetes services provided at the surgery[14].

The fifth objective of diabetes care relates to the reduction of the impact of diabetes on patients. In this study we have attempted to capture this element by assessing the patient perceived quality of services delivery. A multi-item question was developed to capture the elements of primary diabetes care that can impact upon the quality of patients' lives. These elements include easiness and convenience of access to services; clinicians' efforts to involve the patients in the decisions of care; clinicians' time to listen and talk to the patients; the support clinicians give to the patients; consistency of advice given and usefulness of education materials given[15]. For the specific questions used, please refer to the patient questionnaire (Appendix B, section C).

---

[14] This measure was estimated by calculating the average of the proportions of patients who answered 'Always' or 'Most times' to items a-m of Section C of the questionnaire (Appendix B).

[15] Items regarding the comfort of the waiting areas in the surgery and other logistic aspects have not been included because they were not seen as having a significant impact on patients' Quality of life.

6.6.5.8 Patient Confidence

Measure: Proportion of patients who report feeling 'Very Confident' or 'Confident' in taking good control over their diabetes.

In order to be able to actively control their diabetes, patients need to feel confident regarding the usefulness of continuous treatment and lifestyle changes. In a chronic condition, such as diabetes, where the effects of good control are only visible in the long term, having full confidence in being able to control the condition is essential. The role of the health care professionals is very important in building this confidence. Therefore, we believe that an assessment of how confident patients feel in taking good control of their diabetes is essential to evaluate the quality of care provided. The data for this outcome was extracted from the answers to question 3 of section B of the patient questionnaire (Appendix B).

6.7 The Use of Proportions as Outputs in DEA

In our models of efficiency and effectiveness, we have used proportions as outputs, as it was believed that these were the measures that better represented the objectives related to diabetes management. We must stress that the proportion of patients with BP under control is more important than the actual number of patients with BP under control. By using a proportion measure we simultaneously account for service delivery and need. Proportion measures are also the measures used by the Government both in the definition of the NSF PIs and in the performance related pay targets stipulated in the new GP contract. Therefore, it was believed that it is important to use proportions in order to evaluate the performance of the surgeries in these terms and to evaluate the impact on costs of using this type of measures.

However, as pointed out by Fernandez-Castro and Smith (1994), using proportions in DEA can lead to incorrect inferences. The problem is that DEA calculates an approximate weighted average with the proportions, which can be different from the correct weighted average that would be obtained if the ratios were used. However, as suggested by Salinas-Jimenez and Smith (1996), this approximation is often used and it does not lead to serious problems in most studies.

In our case study, we also ran the models using volume measures and the number of surgeries deemed efficient and cost-effective was in fact smaller. This results from the fact that a DEA model that uses proportions as outputs implicitly has more variables than a DEA model that uses volume measures. For example, if the proportion of diabetics with BP under control is used as an output, this is equivalent to using a model where the number of diabetics has been used as a non-controllable input and the number of diabetics with BP under control has been used as an output, and where the weights of these two variables have been restricted to be the same. Whilst the use of volume measures might be considered advantageous given the greater discrimination that it provides, we believe that the use of proportions provides fairer estimates of efficiency and cost effectiveness by taking into account the level of need in each surgery.

However, when the outputs are represented in proportions, we need to represent the inputs as the average amount of resources spent per patient in order to establish that an increase in the inputs should lead to an increase in the outputs. Once this has been done, we can use the CCR model to compare the efficiencies of different units (Cooper, Seiford and Tone 1999).

One further concern regarding the use of proportions as outputs relates to the possibility of obtaining meaningless targets (projections to an output level beyond 100%), if an output oriented model is used. In our case study, whenever proportions were used as outputs, we used input oriented models, which means that meaningful targets can be calculated. Furthermore, as discussed by Dyson et al. (2002), it is important not to combine measures of volume with proportion measures in the same model, as this leads to a break down in terms of proportionality.

6.8 Conclusion

In this chapter we have provided the background for the case study in primary diabetes care. We have justified the need for a formative evaluation in primary diabetes care. The drawbacks of the newly developed National Service Framework for diabetes were also discussed. We have provided a brief characterisation of diabetes in order to facilitate the discussion of the framework for evaluation in primary diabetes care. The needs, objectives, resources, outputs and outcomes of primary diabetes care were also discussed in some detail, in order to justify the performance assessment models used in the next chapter. We will discuss the data available from the participating GP surgeries and report on the results obtained from the data analyses in chapter 7.

7 Data, Models and Results

7.1 Introduction

In this chapter we present the results from the formative evaluation undertaken for a sample of surgeries in terms of their diabetes service delivery. In Section 7.2 we make a brief characterisation of these surgeries in order to familiarise the reader with the providers under comparison and discuss the data limitations. In Section 7.3, we discus the data used and its limitations. In Section 7.4 we detail the analyses based on the performance assessment framework outlined in Chapter 6. In this respect, we present and discuss the results in terms of equity, efficiency and effectiveness. Section 7.5 concludes this chapter.

7.2 A Brief Characterisation of the Surgeries Under Comparison

In the course of this project, we collected data from 23 surgeries, including data from the patient questionnaire and the practice profile questionnaire. At this stage we have decided to perform a detailed analysis of the results of 14 surgeries only, as we are concerned over the quality of the data extracted from the remaining surgeries. For the 14 surgeries studied we are certain that the data has been extracted using consistent methods, and that the data is of reliable quality. This chapter presents the results for this sample of 14 surgeries. The data extracted for the remaining 9 surgeries was then used at a second stage to evaluate the robustness of the results obtained for the performance of the 14 surgeries studied.

In terms of the surgeries included in the analyses, 12 belong to South Peterborough PCT, one belongs to Melton and Rutland PCT and the other belongs to Heartlands PCT.

To maintain anonymity, the surgeries have been given an identification number. Table 7.1 provides an overview of these surgeries regarding their location and size.

Table 7.1: Characteristics of the surgeries

Surgeries	Number of registered patients	Number of diabetics	Location	Total number of GPs (WTE)	Average list size per WTE GP	Total number of Nurses (WTE)	Diabetes clinics
1	11,506	426 (3.70%)	Mixed	5.5	2,092	3.075	GP/Nurse-led twice weekly
2	12,843	346 (2.69%)	Mixed	7.75^{α}	$1,675^{\omega}$	3.6	Nurse-led twice weekly
3	8,654	211 (3.44%)	Suburban	5.5	1,573	3	Nurse-led five times a week
4	10,352	286 (2.76%)	Rural	5.75	1,800	1.5	Nurse-led twice weekly
5	6,958	214 (3.08%)	Rural	3.5^{α}	1,988	3	Nurse-led monthly
6	13,722	461 (3.36%)	Mixed	7.75^{α}	$1,771^{\omega}$	6	GP/Nurse-led weekly
7	5,760	145 (2.52%)	Rural	3^{α}	$1,920^{\omega}$	3	No diabetes clinic
8	3,869	114 (2.95%)	Inner city	2	1,935	1	GP/Nurse-led weekly
9	8,758	239 (2.73%)	Mixed	4^{ε}	2,190	1.5	No diabetes clinic
11	1,334	25 (1.87%)	Rural	1	1,334	0.875	Nurse-led monthly
12	2,045	51 (2.49%)	Suburban	1.5625	1,309	1	Nurse-led weekly
13	3,849	88 (2.29%)	Inner city	1^{ε}	3,849	1.13	No diabetes clinic
23	16,528	471 (2.85%)	Suburban	7^{β}	$2,361^{\omega}$	3.5	GP/Nurse-led daily
24	34,340	972 (2.83%)	Mixed	16.5^{γ}	$2,081^{\omega}$	15	GP/Nurse-led daily

Key

WTE	Whole Time Equivalent
α	One GP trainee
β	Two GP trainees
γ	Three GP trainees
ε	One vacancy
ω	Excluding GP trainees

Most surgeries operate from a single location. However, Surgery 5 operates from three separate locations; Surgery 6 and Surgery 7 operate from two locations. After discussions with the Practice Managers at each one of these three surgeries, it was decided that it would not be relevant to separate the analyses for each one of the locations, as they were said to operate as a whole.

From Table 7.1, we note that the 14 surgeries included in this study vary considerably in size with the smallest surgery (Surgery 11) managing only 1,334 patients, and the largest surgery (Surgery 24) managing 34,340 patients. Furthermore, they vary in terms of the average number of patients per GP. Surgery 11, again, has the smallest list per GP (1,334 patients) and Surgery 13, currently with a GP vacancy managing an average of 3,849 patients per GP. Arguably, this would compromise homogeneity if the objective were one of classification and target setting. However, given that the purpose of this study is *exploratory and formative*, we believe that this may contribute to the discussion regarding the association between list size and the quality of the services provided. Wilkin et al. (1987) in their study of General Practice in Manchester concluded that the ideal list size appears to be between 2,000 and 2,500 patients per WTE GP, and found no evidence that that a list size smaller than 2,000 patients brings additional benefits to the patients.

We can also notice that the diabetes prevalence varies from surgery to surgery. This is partly explained as a result of population characteristics, but it may also be a result of different screening policies. Furthermore, the number of diabetics registered with each surgery varies from 25 to 972, which indicates that different surgeries may have different incentives to implement an effective diabetes management system. One of the possibilities introduced by the new GP contract is that surgeries can 'opt out' of diabetes provision if they feel that its management is not cost-effective (British Medical

Association 2003). In this case, the PCTs will have to provide diabetes services to the patients. Despite the different number of diabetics, all 14 surgeries provide diabetes services to their patients and all of them have a formal lead person for diabetes (a person with a special interest in diabetes). Furthermore, all of the surgeries wanted to be included in the study in order to evaluate their performance in relation to the others.

This leads us to the next issue. It is important to repeat that all 14 surgeries agreed to take part in this research project and were involved in the development of the framework for evaluation and the development of the patient questionnaire. A long process of consultation took place with each surgery prior to the data extraction and the patient survey. In this respect, all these 14 surgeries volunteered to take part in this research study. We feel that this is very important in order to obtain meaningful results, which may be of use in practice. However, considering this process of self-selection, it is possible that the results presented here portray a biased picture in terms of equity, efficiency and effectiveness in primary diabetes care delivery in England. It is natural to expect that these 14 surgeries are already trying to do their best and are all engaged in improving their diabetes care. Our intention is to use the results from this sample to assess the potential of DEA as a tool for formative evaluation in primary heath care; and to identify some of the structures and mechanisms behind successful performance *within* this small sample. However, when we consider the fact that a small self-selected sample was analysed, care needs to be taken in terms of the interpretation of the results and of its generalisation.

As previously discussed, the data collection was the most difficult part in this case study. Firstly, most of the data had never been extracted before. Several discussions with the information facilitators of each PCT took place in order to identify the variables for which there would be reliable data. Extraction queries were then written in order to extract data in a comparable format from all the surgeries. Secondly, this research project coincided with a very busy period both to the PCTs and to the surgeries; as a result, the data collection process was longer than expected. The data used in this case study is presented in Appendix F. Further to the discussion in Chapter 6, there are some specific observations about the data quality that we should make before proceeding to the data analysis.

7.3 The Data and its Limitations

The prescribing data was obtained from the Prescribing Price Authority and the data regarding the number of consultations was extracted from the surgeries' computer register systems. When we consider that all the surgeries included in the analyses operate a computerised appointments system, the data for these inputs should be of reliable quality.

The data for the outputs and the outcomes was extracted from the surgeries' clinical systems, using for each variable all the codes in use at each surgery. For example, with regards to the main output, which is the 'proportion of patients who have had a full diabetes annual review', an effort was made to identify all the codes used in each surgery for this procedure. Despite the fact that there is a standard code for this procedure, this code is not used by all professionals in all surgeries. Some surgeries have created their own codes. Furthermore, it is conceivable that in certain cases, the procedure is carried out but it is not recorded in the computer. Naturally, the data extraction cannot identify cases where procedures were not recorded or where the procedures were wrongly coded[16]. For this reason, the results may be bias against surgeries that are not recording or correctly coding all their activities.

We are convinced that, within our sample, the number of cases that have not been recorded or correctly coded, should be low, because the surgeries have undergone a long period of training regarding clinical coding and recording. This is one of the reasons why we are working with a small sample. Surgeries where the electronic recording was poor had to be excluded. It is expected that with the implementation of the National Service Framework for diabetes, the standard of data in primary diabetes care will improve significantly across all surgeries in England. This should facilitate future implementation of a process of regular formative evaluation in primary care.

7.4 Performance Results

We will now present the results obtained from our comparison of the 14 surgeries in terms of geographical equity, efficiency and effectiveness. Our DEA measures follow the methodology outlined in chapter 2 and the variables included in the different models follow the discussion carried out in chapter 6.

7.4.1 Equity of Service Utilisation

We are interested in measuring equity of utilisation at an aggregated level, in particular, on a geographical level. We must now clarify how we have measured equity of utilisation.

---

[16] It is interesting to note that most clinical computer systems can lead to coding errors when the 'search' facility is used. For example, if the term 'diabetes review' is typed several different codes are suggested, and it is up to the professional to choose the one to use. Obviously, this can cause some problems in terms of standardization.

In Chapter 4, we defined horizontal equity as the assessment of whether groups of patients with a similar level of need receive consistent treatment. Three ways to assess the level of treatment received were discussed: (1) level of resources used, (2) level of services utilisation and (3) level of outcomes achieved. To assess equity based on the level of resources used is problematic because it is influenced by the level of efficiency of each provider. To assess equity based on the level of outcomes achievement requires careful consideration given the strong impact of patients' characteristics on the outcomes. For these reasons, we investigate whether diabetic patients registered with different surgeries show, on average, similar utilisation rates of the services they equally need. We assume that there are a number of services that are essential in diabetes care and that all diabetic patients have an equal need to receive those services. According to British Diabetic Association (2001) there are seven essential services that all diabetics should receive at least once a year (please refer to Model 7.1 below).

We used the patient questionnaire to obtain this data because the surgery records were incomplete regarding the procedures or tests that had been undertaken outside the surgery, as a result of poor coordination with other local providers. We have asked the representative sample of patients from each surgery to indicate the number of times they had certain services over the previous 12 months (please refer to Section 1 of the questionnaire in Appendix B).

Model 7.1: Variables used for the measurement of Equity

Estimated proportion of diabetics who have had these essential services in the previous 12 months

(1) A detailed eye examination, with dilated pupils
(2) A detailed examination of feet and legs
(3) A review of the blood pressure (BP)
(4) A review of blood cholesterol
(5) A review of the blood glucose by means of a HbA1c test
(6) A review of the kidney function, by urine and blood tests to check for protein
(7) A review of the medication, diet and any other relevant matters, with a doctor or a nurse

For an assessment of equity, we are not concerned whether the patients have had the essential diabetes services at the surgery or outside the surgery. Our objective is to assess whether particular groups of patients have systematically not received certain services, for reasons outside their control. This leads us to an essential point in the assessment of equity of utilisation: the fact that different patients may have different propensities to utilise health services. Inequality in utilisation rates does not necessarily mean inequality in access. To account for this problem, we have included an open question in the questionnaire, asking the patients to justify why they have not received particular services. We hope that patients would acknowledge when it was their decision not to attend a consultation or test offered. Examples where the patient acknowledges being offered a service and refusing to take it were excluded from the assessment of equity. For example, we have found that many patients who did not have any consultation with a dietician provided a suitable justification. These patients state that despite being offered an appointment with a dietician, they refused to take it, because they did not find such consultation necessary or useful. For this reason, we decided not to include 'a consultation with a dietician' in the measurement of equity.

To conclude, our measure of service utilisation for the measurement of equity is defined as the proportion of diabetics within each surgery who report having received *all* the essential diabetes services during the 12-month period of this study. This proportion was analysed across our sample of surgeries for the whole set of diabetics. The list of essential services considered is presented in Model 7.1 and the results of this assessment are presented in Table 7.2 below.

Table 7.2: Equity results

Results	Equity	Equity	Equity
Surgeries	**Model 7.1**	**Model 7.1**	**Model 7.1**
	Mean proportion of diabetics who received all services	95% Confidence interval lower limit	95% Confidence interval upper limit
1	84.73%	79.60%	89.86%
2	88.98%	84.39%	93.58%
3	84.88%	79.04%	90.72%
4	90.20%	86.98%	93.41%
5	78.64%	72.93%	84.35%
6	80.54%	75.30%	85.77%
7	72.00%	64.92%	79.08%
8	75.00%	66.28%	83.72%
9	71.56%	65.30%	77.82%
11	64.29%	47.29%	81.28%
12	59.09%	43.44%	74.74%
13	78.38%	68.22%	88.53%
23	77.50%	71.04%	83.96%
24	67.89%	62.43%	73.35%
Average	**76.69%**	**69.08%**	**84.30%**

The average proportion of patients who have received all the essential diabetes services across the 14 surgeries is 76.69%. Therefore, we can conclude that, on average, 23% of the patients have not received some of the essential services in the 12-month period of this study. There may be several reasons for this:

(1) random variability in the data, given that only a sample of patients was surveyed;

(2) patients may have forgotten about a particular check they had, or may have been confused about when they had that check. However, it is reasonable to expect that on average, surgeries will be equally penalised by inaccuracies of patients' memories, and therefore a relative comparison is meaningful;

(3) certain surgeries operate a late 'call and re-call system' for the essential diabetes services, which means that some patients are not reviewed within the recommended 12-month period;

(4) certain surgeries have relatively poor coordination with the local hospital and local opticians, and this means that a proportion of their patients have not had an annual eye check, for example;

(5) certain surgeries have not invited some of their diabetic patients for some of the annual checks.

Once all these factors have been considered, if, for example, one surgery presents a significantly lower proportion of patients who have not received all the essential services, this can be an indication of relatively 'poorer access' in that surgery. Surgery 12 presents the lowest score (59.09%). This score is significantly lower than the average at the 95% confidence level. However, Surgery 4 presents the highest score in terms of equity, given that more than 90% of its patients report having received all the essential services in the 12 month-period of this study. This score is significantly higher than the average at the 95% confidence level. Therefore, we can state that a diabetic patient registered with Surgery 12 has a significantly lower probability of receiving all the essential diabetes services when compared to a patient registered with Surgery 4. Further investigation regarding which particular service is being missed at each surgery should lead towards corrective action.

7.4.2 Technical Efficiency

Technical efficiency refers to the optimal use of resources in the delivery of services. The model used for technical efficiency is Model 7.2 below. We used three inputs and three outputs to measure technical efficiency. It must be remembered that we only have a small sample of surgeries to compare and, as a result, we had to reduce the number of variables, in order to obtain some discrimination between the surgeries. Instead of using the quantities for all the individual medication items prescribed, we have used an average number of medication items prescribed to each diabetic. We previously defined that an item of medication is the average dosage necessary for 28 days of treatment. The two other inputs used were the WTE GP and the WTE nurse per diabetic. The first output represents the proportion of patients who have had a diabetes annual review in the previous 12 months. The second output measures the proportion of patients who are receiving anti-hypertensive medication (ACEI). The third output measures the proportion of patients under regular treatment with lipid-lowering medication (statins).

An input-oriented model was used in order to calculate the minimum quantities of resources that allow the current delivery of diabetes services. We have assumed constant returns to scale as Model 7.2 only uses variable inputs that should be proportional to the number of registered diabetic patients. Any diseconomies of scale that might have existed have been reduced by the fact that the PCT contracts specialist professionals, which go to each surgery to provide extra diabetes care, if needed. Therefore, as we could not find any evidence of economies or diseconomies of scale, we have decided to use a CRS model.

We have decided not to introduce any weight restrictions in the models because we want to decompose overall efficiency into its technical and allocative components. For this reason the input weights should be unrestricted so that the input mix effect is captured by comparing the results of a model with total flexibility of input weights and a model which uses the input unit prices as the weights. The software used to run the DEA models was *Onfront 2* developed by the Economic Measurement and Quality in Lund Corporation (1998-2000) and *EMS 1.3* developed by Holger Scheel. Table 7.3 presents the results of technical efficiency.

Model 7.2: Variables used for the measurement of Technical Efficiency – input oriented

Inputs	Outputs
(1) WTE GP per diabetic (2) WTE Nurse per diabetic (3) Average number of medication items prescribed per diabetic	(1) Proportion of diabetics who have had a complete diabetes Annual Review (2) Proportion of diabetics on targeted ACEI (3) Proportion of diabetics on targeted statins

Table 7.3: Technical efficiency results

Results	Technical Efficiency	Peers
Surgeries	**Model B**	**Model B**
	Fi(y,x) | CRS)	**Fi(y,x) | CRS)**
1	85.52%	6 (0.10) 9 (0.14) 23 (0.66)
2	100.00%	2
3	83.00%	6 (0.18) 9 (0.03) 23 (0.62)
4	98.29%	6 (0.34) 8 (0.39) 23 (0.34)
5	86.72%	2 (0.22) 6 (0.53) 23 (0.04)
6	100.00%	6
7	93.81%	2 (0.34) 6 (0.47) 23 (0.21)
8	100.00%	1
9	100.00%	3
11	100.00%	0
12	71.20%	9 (0.53) 23 (0.25)
13	100.00%	1
23	100.00%	7
24	86.09%	6 (0.47) 13 (0.14) 23 (0.38)
Average	**93.19%**	

From Table 7.3 we can see that Model 7.2 results in 7 out of the 14 surgeries being classified as technically efficient. The average efficiency score is 93.19%. For the efficient surgeries it is of interest to investigate the number of times that each surgery appears as a peer to non-efficient surgeries. This provides us with an indication regarding the comparison group that each surgery used to classify itself, and an indication regarding the robustness of its score.

For example, Surgery 23 is peer to 7 surgeries, Surgery 6 is peer to 6 surgeries and Surgery 9 is peer to 3 surgeries. From another side, Surgery 11 is only peer to itself. This means that Surgery 11 is able to appear efficient by choosing a unique weight structure. In particular, in order to appear efficient, Surgery 11 puts most of its input weights on the WTE nurse per patient, and all of its output weights on the proportion of patients under regular treatment with ACEI. When we consider that no other surgery benefits from this weight structure, Surgery 11 does not have a comparison group and appears efficient as a result. Surgery 11 can then be defined as a *self-evaluator* DMU.

Surgery 12 presents the lowest score in terms of technical efficiency. An analysis of the optimal values for the intensity variables reveals that Surgery 9 and Surgery 23 are the peers for Surgery 12.

This means that a virtual surgery can be created by using the following convex combination:

$$\left(0.53 \times Surgery_{9} + 0.25 \times Surgery_{23}\right).$$

This virtual surgery delivers *at least* the same quantities of outputs as Surgery 12, but uses only 71.20% of the inputs used by Surgery 12. Further investigation reveals that with this reduced set of inputs is possible to achieve greater levels of outputs, as there is a 7.29% slack on the proportion of patients under regular treatment with ACEI; and 7.03% slack on the proportion of patients under regular treatment with statins.

7.4.3 Cost Efficiency - Introducing Information about Input Prices

Some of the surgeries have used a non-economical weight structure in order to increase their efficiency score. For example, Surgery 11 was classified as technically efficient by placing all its input weight on the WTE nurse whilst ignoring the other two inputs. This reflects an input mix that is sub-optimal in terms of allocative efficiency. Furthermore, it leads to unrealistic marginal rates of input substitution. When we consider that we have estimates for the input prices, it is beneficial to use this information in order to calculate cost efficiency. This type of efficiency aims to minimise the costs, whilst maintaining the current level of outputs. Cost efficiency is the most demanding of all efficiency measures. It requires an optimisation in terms of quantities of inputs used to produce the outputs, but also it requires that the input mix is optimal given the current input prices. Once an estimate of cost or overall efficiency is obtained it is of interest to decompose it into its constituent components, as discussed in chapter 2.

To calculate cost efficiency we have used Model 7.2, together with the following labour unit costs: £86 per hour of WTE GP; £23 per hour of WTE nurse. The average unit cost for the medication varies between surgeries, and the figures are presented in Appendix F (Table F5).

7.4.3.1 Decomposition of Cost Efficiency

We have decomposed our measure of cost efficiency into allocative efficiency and pure technical efficiency, in order to obtain some insights regarding the sources of inefficiency in each surgery.

Allocative efficiency is calculated by comparing the score of cost efficiency with the score of technical efficiency. It is worth noting that if a certain surgery is not using the most adequate input mix, its score of cost efficiency will be inferior to its score of technical efficiency.

Table 7.4 presents the results for cost efficiency, allocative efficiency and technical efficiency. The mean level of cost efficiency is 85.28%, which indicates that on average savings of 14.72% could be made on the costs, whilst maintaining the current level of service delivery. Further investigation of the optimal input values for each surgery will pin point which inputs need to be reduced and which inputs need to be increased in order to obtain cost efficiency. We have two surgeries that are cost efficient: Surgery 2 and Surgery 23. Once again, surgery 12 presents the lowest efficiency score. In order to understand the sources of the inefficiency in each one of the inefficient surgeries, we need to analyse the results from the decomposition into allocative and technical efficiency.

Table 7.4: Decomposition of cost efficiency

Results	Cost efficiency	Allocative efficiency	Technical efficiency
Surgeries	Model B	Model B	Model B
	Oi(y,x,w) | CRS)	Ai(y,x,w) | CRS)	Fi(y,x) | CRS)
1	80.64%	94.29%	85.52%
2	100.00%	100.00%	100.00%
3	74.38%	89.61%	83.00%
4	88.75%	90.29%	98.29%
5	76.77%	88.53%	86.72%
6	89.18%	89.18%	100.00%
7	84.09%	89.64%	93.81%
8	94.70%	94.70%	100.00%
9	94.86%	94.86%	100.00%
11	71.64%	71.64%	100.00%
12	68.63%	96.40%	71.19%
13	91.69%	91.69%	100.00%
23	100.00%	100.00%	100.00%
24	78.55%	91.24%	86.09%
Average	**85.28%**	**91.51%**	**93.19%**

7.4.3.2 Allocative Efficiency and Input Mix

In terms of allocative efficiency, which strictly measures how the current input mix fits the input prices, we have an average score of 91.51%. Surgery 12 is no longer the worst performer using this criterion, presenting a score of 96.40%. This means that its low cost efficiency results primarily from technical efficiency. Its technical efficiency score is 71.20%. After radially reducing all its inputs to 71.20% of their current value, Surgery 12 removes most of its inefficiency.

On the contrary, Surgery 11, which was classified as technically efficient, presents a score of 71.64% in terms of allocative efficiency. As we have suggested before, its input mix does not fit the current input prices. We can recall that this surgery was a *self-evaluator* and was classified as efficient by assigning all its input weights to WTE nurse, which is economically unreasonable. We can further investigate the problem with the input structure of Surgery 11 by comparing its current input mix with its optimal input mix. Table 7.5 presents the values for the current input quantities for each surgery. For example, Surgery 11's current ratio of WTE GP to WTE nurse is 2.53. This is a very costly input mix, given that one hour of WTE GP is nearly 4 times more expensive than that of WTE nurse. To further this situation, Surgery 11 presents the highest figure in terms of the number of medication items prescribed per patient. This surgery is classified as cost inefficient, because its costly use of resources is not compensated by a high enough level of service delivery.

We can recall that Surgery 11 is the smallest surgery working with a single GP. It has the smallest number of registered patients (1,334), which means that this GP has more time to spend with its patients than GPs in other surgeries. Whilst this is a positive aspect from the patients' point of view, there is certainly an opportunity cost associated with its relatively small list size. It is of interest to investigate whether this extra time brings additional benefits in terms of outcomes achievement and we will investigate that later.

Surgery 23 is cost efficient, despite having a high ratio of WTE GP to WTE Nurse. This is because Surgery 23 prescribes a lower than average number of medication items, and is able to deliver a large enough number of services to compensate for the expensive labour mix chosen. For example, this surgery presents the highest figure in terms of the proportion of patients that have had an annual review in the 12-month period (83.23%).

Table 7.5: Current and optimal input mix

Results	Current input quantities			Optimal input quantities – Model B		
Surgeries	WTE GP Hours per patient	WTE Nurse Hours per patient	Number of medication items prescribed per patient (average)	WTE GP Hours per patient	WTE Nurse Hours per patient	Number of medication items prescribed per patient (average)
1	1.85	1.18	21.32	1.63	0.76	16.84
2	1.77	2.08	22.28	1.77	2.08	22.28
3	1.74	2.16	20.67	1.49	0.75	15.55
4	1.65	1.70	23.59	1.77	1.30	19.78
5	1.32	1.61	21.80	1.42	0.85	15.33
6	1.27	1.5	24.81	1.84	0.86	19.06
7	1.70	1.7	24.86	1.66	1.54	19.55
8	1.43	1.82	21.06	1.59	1.55	19.02
9	1.53	1.23	20.53	1.79	0.84	18.48
11	2.23	0.88	31.17	1.77	1.52	20.51
12	1.80	1.27	22.12	1.42	0.66	14.66
13	1.25	1.59	17.43	1.49	0.70	15.37
23	1.88	0.88	19.45	1.88	0.88	19.45
24	1.73	2.05	25.01	1.79	0.84	18.54
Average	**1.65**	**1.55**	**22.58**	**1.67**	**1.08**	**18.17**
Min.	**1.25**	**0.88**	**17.43**	**1.42**	**0.66**	**15.33**
Max.	**2.23**	**2.16**	**31.17**	**1.88**	**2.08**	**22.28**

7.3.3.3 Including Information about Consultations with Specialists in the Measurement of Efficiency

In all the models used until now, we have ignored the number of consultations with diabetes related specialists. We are referring to the number of consultations with an ophthalmologist, a diabetologist and a dietician, which should be included in the model in order to make a fairer evaluation of a surgery's efficiency. We have excluded this information in most of our performance models, as the data extracted from the surgeries for these variables is not reliable. This type of data tends to be kept in the form of letters, which are not always entered in the computer. The extraction of this data is therefore very difficult.

The Diabetes Services Questionnaire was designed to obtain estimates for this data. We asked the patients how many times in the previous 12 months they had had each type of consultation. Using their responses we were able to estimate data for the average number of consultations with an ophthalmologist, a diabetologist and a dietician. We must remember that this data was estimated using the responses from the patient questionnaire, and as a result we need to take heed regarding its interpretation. Model 7.3 includes the estimates for the average number of consultations with an ophthalmologist, a diabetologist and a dietician. The efficiency results from this model are presented in Table 7.6.

The results for technical efficiency increase significantly after the inclusion of this information. Three new inputs are included, and considering that no weight restrictions have been introduced, each surgery now has three new variables to distinguish itself. Only three surgeries are classified as technically inefficient after the inclusion of data about consultations with the specialists. These are surgeries 4, 5 and 12.

Model 7.3: Variables used for the measurement of cost efficiency (including estimates for consultations with specialists)

Inputs	Outputs
(1) WTE GP per registered diabetic (2) WTE Nurse per registered diabetic (3) Average number of medication items prescribed per diabetic (4) Estimated average number of Ophthalmologist consultations per diabetic (5) Estimated average number of Diabetologist consultations per diabetic (6) Estimated average number of dietician consultations per diabetic	(1) Proportion of diabetics who have had a complete diabetes annual review (2) Proportion of diabetics on targeted ACEI (3) Proportion of diabetics on targeted statins

We have also calculated cost efficiency in order to take into account the unit prices of these inputs. The average cost efficiency score, after the inclusion of information about the consultations with specialists is 86.24%, which is only slightly higher than the average cost efficiency score of Model 7.2. We note that some surgeries benefit from the inclusion of this data, whilst others see their cost efficiency scores drop. For example, Surgery 4 sees its cost efficiency score drop from 88.75% to 87.27%. However, Surgery 6 is the surgery that most benefits from the inclusion of this information. Surgery 6 has the lowest referral rates to a dietician and a diabetologist and, for this reason, sees its cost efficiency score increase from 89.18% to 95.63%.

We know that most surgeries appear to be over-prescribing (Table 7.5), when we consider their current service delivery. For example, Surgery 24 currently prescribes an average of 25.01 medication items per patient, whilst its optimal amount of medication items prescribed should be 18.54. If we wanted to identify exactly which items of medication ought to be reduced, a different model discriminating between the nine types of medication should be used (Model 7.4).

Table 7.6: Efficiency results including costs with specialists

Results	Cost Efficiency	Allocative Efficiency	Technical Efficiency
Surgeries	Model C	Model C	Model C
	Oi(y,x,w) | CRS)	Ai(y,x,w) | CRS)	Fi(y,x | CRS)
1	80.63%	80.63%	100.00%
2	100.00%	100.00%	100.00%
3	73.61%	73.61%	100.00%
4	87.27%	88.15%	99.00%
5	76.66%	88.12%	87.00%
6	95.63%	95.63%	100.00%
7	86.29%	86.29%	100.00%
8	94.69%	94.69%	100.00%
9	96.24%	96.24%	100.00%
11	72.02%	72.02%	100.00%
12	70.61%	82.10%	86.00%
13	93.52%	93.52%	100.00%
23	100.00%	100.00%	100.00%
24	80.18%	80.18%	100.00%
Average	**86.24%**	**87.94%**	**98.00%**

Model 7.4: Variables used for the measurement of cost efficiency, using detailed measures for the medication prescribed (input oriented)

Inputs	Outputs
(1) WTE GP per registered diabetic (2) WTE Nurse per registered diabetic (3) Average number of Biguanides prescribed per diabetic (4) Average number of Sulphonylureas prescribed per diabetic (5) Average number of other anti-diabetic drugs prescribed per diabetic (6) Average number of items of screening and monitoring prescribed per diabetic (7) Average number of hypoglycaemic drugs prescribed per diabetic (8) Average number of short-acting insulin items prescribed per diabetic (9) Average number of intermediate/long-acting insulin items prescribed per diabetic (10) Number of ACEI items prescribed per diabetic (11) Number of Statins prescribed per diabetic (12) Estimated average number of Ophthalmologist consultations per diabetic (13) Estimated average number of Diabetologist consultations per diabetic (14) Estimated average number of Dietician consultations per diabetic	(1) Proportion of diabetics who have had a complete diabetes annual review (2) Proportion of diabetics on targeted ACEI (3) Proportion of diabetics on targeted statins

Given that a large number of variables are used in this model, we have not calculated technical efficiency using Model 7.4. Nevertheless, by applying the individual prices to each one of the inputs, the minimum feasible cost can be estimated and we can compare the surgeries in terms of their cost efficiency. In this respect, Model 7.4 is our most complete model of cost efficiency, as it includes the costs with the specialists and discriminates between the individual types of medication. The results for this model are presented below in Table 7.7.

Using the results of Model 7.4, we note that the actual average annual unit cost per diabetic varies from £535.5 to £831.35. When we consider the level of service delivery, our cost efficiency comparison suggests that some of this variation appears to be unjustified. However, we need to analyse effectiveness in order to understand the effects of different cost levels on outcomes achievement, as detailed in the next section.

Table 7.7: Cost efficiency results, based on the decomposition of the medication

Results	Actual average cost per diabetic (£)	Minimum average cost per diabetic (£)	Cost efficiency
Surgeries	**Model D**	**Model D**	**Model D**
	WX – Actual cost	**C(y, x, w) | CRS) - Minimum cost**	**Oi(y,x,w) | CRS)**
1	631.77	529.63	83.83%
2	657.01	657.01	100.00%
3	641.97	480.32	74.82%
4	711.39	601.96	84.62%
5	599.60	483.13	80.58%
6	580.13	575.15	99.14%
7	668.80	578.90	86.56%
8	592.73	580.01	97.85%
9	589.14	589.14	100.00%
11	831.35	610.52	73.44%
12	648.11	447.44	69.04%
13	535.50	511.36	95.49%
23	622.44	622.44	100.00%
24	720.13	576.20	80.01%
Average	**645.01**	**560.23**	**87.53%**

7.4.4 Clinical Effectiveness

Model 7.5 uses the outputs from the efficiency models as inputs and uses three clinical outcomes as outputs. This model does not consider the delivery costs; it evaluates the impact of the services delivered on the clinical well being of the patients. The three clinical outcomes chosen assess the proportion of patients with the three essential biomedical indicators under control. This model assesses the 'added value' of the health care professionals' time and medication in terms of these three essential clinical outcomes. We have applied an output-orientation to Model 7.5, as the objective is to maximise the outcomes, given the level of services delivered.

Model 7.5: Variables used for the measurement of clinical effectiveness (output oriented)

Inputs	Outcomes
(1) Number of diabetics who have had a complete diabetes annual review	(1) Number of diabetics with BP under control
(2) Number of diabetics on targeted ACEI	(2) Number of diabetics with blood glucose under control
(3) Number of diabetics on targeted statins	(3) Number of diabetics with blood cholesterol under control

Weight Restrictions (WR)

$$-v_1 + v_2 \leq 0$$

$$-v_1 + v_3 \leq 0$$

We have included two weight restrictions in this model in order to guarantee that no surgery would be classified as effective simply because it had the lowest number of patients under regular treatment with statins, for example. We have developed these two weight restrictions based on the identification of technologically realistic trade-offs between the different inputs. The restrictions introduced and their interpretation is presented below.

7.4.5 Weight Restrictions and their Economic Interpretation

As defined in chapter 2, the weight associated with input i is denoted by v_i and the weight associated with output r is denoted by u_r. The general formula for the introduction of homogeneous weight restrictions according to the trade-offs approach is given by the following equation (Podinovski 2002):

$$a_1u_1 + a_2u_2 + ... + a_su_s - b_1v_1 - b_2v_2 - ... - b_mv_m \leq 0.$$

7.4.5.1 Annual Review and Regular Treatment with ACEI

If the number of patients under regular treatment with ACEI (x_2) decreases by one and the number of patients with a full annual review (x_1) increases by one, there should be no impact on the level of outcomes achievement. The justification for this restriction is that the annual review is the most basic service delivered to all diabetics and, per se, is the service that can lead to greater improvements in the outcomes. Regular treatment with medication can only lead to outcomes improvement if the patient is also annually reviewed to ascertain if the medication is working as expected. Symbolically, this can be represented by the following trade-off coefficients:

$$(a_1,\ a_2,\ a_3) = (0,\ 0,\ 0) \text{ and } (b_1,\ b_2,\ b_3) = (1,\ -1,\ 0).$$

Applying these coefficients to the general formula results in the following weight restriction:

$$-v_1 + v_2 \leq 0$$

7.4.5.2 Annual Review and Regular Treatment with Statins

If the number of patients under regular treatment with statins (x_3) decreases by one and the number of patients with a full annual review (x_1) increases by one, there should be no impact on the level of outcomes achievement. The same justification applies. Symbolically, this can be represented by the following trade-off coefficients:

$$(a_1,\ a_2,\ a_3) = (0,\ 0,\ 0) \text{ and } (b_1,\ b_2,\ b_3) = (1,\ 0, -1).$$

Applying these coefficients to the general formula results in the following weight restriction:

$$-v_1 + v_3 \leq 0$$

The imposition of weight restrictions cannot improve the effectiveness score. However, it will reduce the score, unless the restriction is redundant for a particular surgery, in such an example it would have no effect on its score.

For this reason, it is very important to evaluate the appropriateness of the restrictions introduced. If the process that transforms inputs into outputs is complex, and we cannot be certain of the type of trade-offs that can occur, it is prudent to allow weight flexibility. We have chosen not to introduce weight restrictions between the outcomes for this reason. The establishment of realistic causal links between the level of services delivered and the achievement of outcomes is tentative. As a result, we chose to allow full flexibility regarding outcome trade-offs in order to determine conservative effectiveness scores. Table 7.8 presents the results of clinical effectiveness.

The average score in terms of clinical effectiveness is 68.63%, which is relatively low. There are a group of surgeries that appear to be suffering from very low clinical effectiveness. It is worth further investigation to determine why surgeries 1, 3, 7, 8 and 13 have clinical effectiveness scores below average. Could it be population demographics, lack of professional training, or a combination of factors? An alternative explanation is that these surgeries might be focusing on other outcomes such as patients' satisfaction at the expense of the achievement of the clinical outcomes. These are important questions that we will try to address.

There are five surgeries that have been deemed clinically effective. There is no evidence that these surgeries could achieve a higher level of outcomes from the services that they deliver. However, some surgeries have been classified as clinically effective by choosing a unique weight structure. For example, Surgery 24 is clinically effective, but is not peer to any other surgery. This is also the largest surgery in the sample and employs a F/T specialist diabetes nurse. Furthermore, the nurse's availability onsite results in a system of very frequent consultations and intensive use of diabetes related medication. This model of care is relatively unique, but nevertheless appears to compensate in terms of clinical effectiveness.

Surgeries 4, 5 and 12, which scored below average in terms of cost efficiency, are clinically effective. Surgery 12 is peer to 9 other surgeries and can therefore be considered one of the role models in terms of clinical effectiveness. However, it is important to note that this is a small surgery working with a single GP, who delivers the majority of the diabetes consultations. The application of this model of care to some of the other surgeries may therefore be difficult. Furthermore, the fact that this surgery presents the lowest score in terms of equity may suggest some caution regarding its use as best practice role model in diabetes care delivery. On the contrary, Surgery 4 is able to deliver effective services, whilst reviewing a very large proportion of the patients (surgery 4 achieved the highest score in terms of equity). Further investigation of the structures and mechanisms used in this surgery can therefore be of significant value for diabetes care delivery.

Table 7.8: Clinical effectiveness results

Results	Clinical Effectiveness	Peers
Surgeries	**Model 7.5 with WR**	**Model 7.5 with WR**
	1/Fo(x,y)	
1	30.26%	5 (0.39) 12 (8.41)
2	66.97%	12 (6.50)
3	46.72%	12 (4.28)
4	100.00%	3
5	100.00%	2
6	100.00%	1
7	44.52%	4 (0.22) 12 (1.29)
8	12.64%	12 (2.09)
9	90.27%	4 (0.51) 12 (2.62)
11	70.59%	12 (0.25)
12	100.00%	9
13	19.25%	12 (1.37)
23	79.60%	4 (1.47) 5 (0.19) 6 (0.04) 12 (1.93)
24	100.00%	0
Average	**68.63%**	

Surgery 23, which was cost efficient, scores 79.60% in terms of clinical effectiveness. We have measured efficiency and clinical effectiveness using a method that implies a trade-off between these two criteria because the outputs in the cost efficiency models are now inputs to the clinical effectiveness models. The professionals present in the workshops believe that there is a trade-off between these measures and therefore the framework had to represent it. Nevertheless, the results suggest that this trade-off is not inescapable. It is possible for a surgery to score very highly in both these criteria. For example, surgery 6 is clinically effective and scored 99.14% in terms of cost efficiency. Whether this balance achieved by surgery 6 comes at a price in terms of patient satisfaction is worth further investigation. We now discuss the evaluation of effectiveness from the patients' perspective, by including patient-centred outcomes.

7.4.6 Patient-focused Effectiveness

The outcomes that have been included in the prior analyses reflect clinical outcomes only. Although these are important types of outcomes to evaluate, there are other outcomes that must be considered. In any study of patient care, the patients' perspective must be given high priority. In order to evaluate the quality of the diabetes services from the patients' perspective; we have defined three patient-focused outcomes. The first outcome aims to measure the effectiveness of patient education and represents the number of patients who report a good understanding about diabetes and its management[17]. Patient understanding represents an important outcome to patients because 99.2% of the patients who answered the patient questionnaire said that having a good understanding of their diabetes and its management is important or very important to them.

[17] Given that several items regarding diabetes management were included in the questionnaire, we calculated the number of patients who answered 'I understand enough' to each item, and then used the average between the results obtained for items a-i. Please refer to Section B of the questionnaire (Appendix B) for the specific items included in this outcome measure.

The second outcome aims to measure the quality of diabetes care, and represents the number of patients who answered 'Always' or 'Most times' regarding the quality of the diabetes services delivered at the surgery[18].

The third outcome aims to measure the extent of patient empowerment, and represents the number of patients that reported being confident or very confident in taking good control of their diabetes. We used an output oriented DEA model (Model 7.6) in order to maximise these patient-focused outcomes. The results obtained are presented below in Table 7.9.

Model 7.6: Variables used for the measurement of Patient-focused Effectiveness (output oriented)

Inputs	Patient centred outcomes
(1) Number of diabetics who answered the diabetes service questionnaire	(1) Average number of diabetics who report that they understand enough about diabetes and its management (2) Average number of diabetics who answered 'Always' or 'Most times' regarding the quality of the diabetes services delivered at the surgery (3) Average number of diabetics who report being 'confident' or 'very confident' in taking good control over their diabetes

[18] An average measure was calculated regarding the number of patients who answered 'Always' or 'Most times' for the items included in Section C of the questionnaire (Appendix B).

Table 7.9: Patient-focused Effectiveness results

Results	Patient-focused Effectiveness
Surgeries	Model 7.6
	1/Fo(x,y)
1	95.24%
2	97.80%
3	95.48%
4	100.00%
5	96.31%
6	92.54%
7	98.63%
8	92.33%
9	91.26%
11	100.00%
12	97.33%
13	94.44%
23	92.35%
24	95.80%
Average	**95.68%**

In terms of patient-focused outcomes, we only have two surgeries that are classified as effective, which are surgeries 4 and 11. Surgery 4 is a medium size surgery, which employs a P/T specialist nurse, who performs all the annual reviews and undertakes most of the diabetic consultations. Surgery 11 has a single GP who undertakes most of the diabetes consultations.

The scores for the remaining surgeries are generally high, but surgeries 6, 9 and 23, which were classified as efficient, score below average in terms of patient-focused effectiveness.

It is important to emphasise that some of the outcomes included in this model may be highly influenced by population demographics. In particular, surgeries located in highly deprived areas may face more difficulties in achieving a good level of patient understanding and patient confidence (in chapter 9 we return to this issue and analyse the impact of environmental variables on the performance results). We will now compare the level of resources expenditure with the level of outcomes achievement across the surgeries.

7.4.7 Cost effectiveness

Whilst clinical effectiveness measures whether a surgery is doing the 'right things', cost efficiency measures whether a surgery is doing things in the 'right way'. In this respect, considering the scarcity of resources, it is essential to identify models of care that achieve the best outcomes at the lowest possible cost. In order to evaluate this, we turn to Models 7.7 and 7.8. These two models use the resources as the inputs (the same inputs as used to measure cost efficiency in model 7.4) and use the different outcome measures as the outputs. Model 7.7 takes into account the three clinical outcomes and Model 7.8 takes into account the patient-focused outcomes.

Model 7.7: Variables used for the measurement of Cost Effectiveness – clinical focus (input oriented)

Inputs	Clinical outcomes
(1) WTE GP per registered diabetic (2) WTE Nurse per registered diabetic (3) Average number of Biguanides prescribed per diabetic (4) Average number of Sulphonylureas prescribed per diabetic (5) Average number of other anti-diabetic drugs prescribed per diabetic (6) Average number of items of screening and monitoring prescribed per diabetic (7) Average number of hypoglycaemic drugs prescribed per diabetic (8) Average number of short-acting insulin items prescribed per diabetic (9) Average number of intermediate/long-acting insulin items prescribed per diabetic (10) Number of ACEI items prescribed per diabetic (11) Number of Statins prescribed per diabetic (12) Estimated average number of Ophthalmologist consultations per diabetic (13) Estimated average number of Diabetologist consultations per diabetic (14) Estimated average number of Dietician consultations per diabetic	(1) Proportion of diabetics with BP under control (2) Proportion of diabetics with blood glucose under control (3) Proportion of diabetics with blood cholesterol under control

Table 7.10 presents the results for cost effectiveness using these different models. Let us focus firstly on the results from Model 7.7. Surgery 6 is cost effective, as previous analyses have suggested (this surgery is both cost efficient and clinically effective). Surgeries 9 and 23 are also classified as cost effective, despite the fact that they were not classified as clinically effective.

Model 7.8: Variables used for the measurement of Cost Effectiveness – patient focus (input oriented)

Inputs	Patient centred outcomes
(1) WTE GP per registered diabetic (2) WTE Nurse per registered diabetic (3) Average number of Biguanides prescribed per diabetic (4) Average number of Sulphonylureas prescribed per diabetic (5) Average number of other anti-diabetic drugs prescribed per diabetic (6) Average number of items of screening and monitoring prescribed per diabetic (7) Average number of hypoglycaemic drugs prescribed per diabetic (8) Average number of short-acting insulin items prescribed per diabetic (9) Average number of intermediate/long-acting insulin items prescribed per diabetic (10) Number of ACEI items prescribed per diabetic (11) Number of Statins prescribed per diabetic (12) Estimated average number of Ophthalmologist consultations per diabetic (13) Estimated average number of Diabetologist consultations per diabetic (14) Estimated average number of Dietician consultations per diabetic	(1) Average proportion of diabetics who report that they understand enough about diabetes and its management (2) Average proportion of diabetics who answered 'Always' or 'Most times' regarding the quality of the diabetes services delivered at the surgery (3) Average proportion of diabetics who report being 'confident' or 'very confident' in taking good control over their diabetes

The results from Model 7.7 suggest that the five surgeries that scored poorly in terms of clinical effectiveness, continue to score poorly in terms of cost effectiveness, which means that their low outcome achievement is not explained by low use of resources. They are using resources, but, for some reason, they are not able to achieve a good level of clinical outcomes.

Surgeries 8 and 13 are of particular interest because they scored relatively high in terms of cost efficiency, which means that they are able use their resources to deliver a relatively high quantity of services, but these services fail to have an impact on the patients' clinical outcomes. These are the only two surgeries located in inner-city areas, which might partly explain their difficulties in achieving good clinical outcomes. Surgery 11 also scores below average in terms of cost effectiveness, because it does not deliver enough services for its current level of expenditure. However, Surgery 11 is located in a pleasant rural village where deprivation is not considered to be an issue.

Table 7.10: Cost Effectiveness results

Results	Cost Effectiveness Clinical Focus	Cost Effectiveness Patient Focus
Surgeries	**Model 7.7**	**Model 7.8**
	Oi(y,x,w) | CRS)	**Oi(y,x,w) | CRS)**
1	31.69%	91.41%
2	53.17%	84.47%
3	41.16%	82.33%
4	96.10%	85.24%
5	83.27%	99.33%
6	100.00%	85.96%
7	30.37%	80.96%
8	11.27%	89.72%
9	100.00%	95.95%
11	23.31%	65.57%
12	91.08%	91.95%
13	16.51%	100.00%
23	100.00%	89.81%
24	86.37%	79.83%
Average	**61.74%**	**87.32%**

Finally, we ran a DEA model (Model 7.8) that aimed to minimise the costs given the level achieved for the three patient-focused outcomes. An analysis of these results shows that the average score is 87.32%. The only surgery that is classified as cost effective under this perspective is Surgery 13. This surgery achieves the highest level of patient confidence, with the lowest level of costs per patient. Surgery 11, on the contrary, presents the lowest score in terms of patient-focused cost effectiveness (65.57%). Despite the fact that this surgery presents relatively high levels of patient-focused outcomes (it scores 100% in terms of patient-focused effectiveness – Model 7.6), once the level of resources used is taken into account, its score drops significantly. This suggests that there are surgeries achieving similar levels of patient-focused outcomes, whilst using fewer resources.

7.5 Conclusion

In this chapter we have discussed the data used for formative evaluation and have presented the results obtained from a comparison of 14 surgeries. The use of DEA has provided interesting insights regarding the decomposition of efficiency and effectiveness measures. Whilst some surgeries are technically inefficient and appear to over-use on all the resources, others surgeries appear to be using a non-optimal mix of resources given their current prices.

The important aspect to remember is that the DEA results are only as good as the model used, and always provide a partial view of performance due to the variables that might have been excluded. One example of this problem relates to the need to include information regarding the level of referrals, given that what appears to be an over-use of in-house resources may actually be compensated by a reduced rate of referrals. In this respect, we emphasise the importance of taking a critical stance when looking at the DEA results.

In the next chapter we analyse the relationships between the different performance measures and look for surgeries that appear to be robust performers, that is, performers that score reasonably well from different perspectives.

8 Analysis of the Results – Looking for Robust Performers

8.1 Introduction

In the previous chapter, we presented the results in terms of equity; efficiency; clinical effectiveness; patient-focused effectiveness and cost effectiveness. In this chapter we analyse the relationships between these different performance criteria in order to evaluate the potential trade-offs between them and in order to identify surgeries that appear to be robust providers. By robust providers, we mean surgeries that score above average in most performance criteria. These are our benchmark surgeries. We find that an analysis of the relationships between the results for the different performance criteria is advantageous both to raise further questions and to point towards surgeries that may benefit from further investigation. Furthermore, it makes explicit the fact that the 'location' of a surgery in terms of performance *changes* with different criteria.

We also analyse the results in terms of robustness. When we consider the fact that a small sample was analysed, it is of interest to investigate the sensitivity of the results to the inclusion of new surgeries. In this respect, we add the data from nine further surgeries to the sample and apply the models to this larger dataset in order to investigate the changes in the pattern of the results. These nine surgeries were the additional surgeries for which we had collected data, but which were not included in the main data analyses because of concerns regarding the quality of the data.

8.2 Trade-offs between the Different Measures of Performance

In this section we analyse the trade-offs between the different performance measures by mapping the performance results of one criterion against those of a different criterion.

We explore the relationship between equity and efficiency, efficiency and clinical effectiveness, equity and clinical effectiveness and clinical effectiveness and patient-focused effectiveness. Table 8.1 presents the Spearman rank correlations between the results for the different performance criteria[19]. We found that the results in terms of cost efficiency are negatively correlated with the results in terms of patient-focused effectiveness. This relationship is significant at the 0.05 level. Apart from this relationship no other significant correlation could be found between the results for the different performance criteria. We will now discuss in detail each one of the hypothesised trade-offs.

Table 8.1: Spearman's rank correlations between the different performance measures

	Equity	Cost efficiency	Clinical effectiveness	Patient-focused effectiveness
Equity	1.000 .	0.302 (0.293)	-0.110 (0.708)	0.059 (0.840)
Cost efficiency	0.302 (0.293)	1.000 .	-0.205 (0.481)	-0.539* (0.047)
Clinical effectiveness	-0.110 (0.708)	-0.205 (0.481)	1.000 .	0.239 (0.412)
Patient-focused effectiveness	0.059 (0.840)	-0.539* (0.047)	0.239 (0.412)	1.000 .

P-Values[20] given in parenthesis

* Correlation is significant at the 0.05 level (2-tailed).

---

[19] The Spearman rank correlation coefficient calculates the correlation between the ranks of the different measures. This coefficient was used because the performance results follow a non-normal distribution.

[20] Probability that the observed relationship between the variables occurred by chance.

8.2.1 Trade-off between Equity and Efficiency

Okun (1975) suggested that the most efficient solution is not the one that leads to the most equitable results. In order to reach a large proportion of patients, higher levels of resources have to be used. There is an additional problem which relates to the fact that in some cases the patients that are the most in need of care are those that are most difficult to reach, and therefore, are those that will lead to greater marginal costs. For this reason the patients that most need the services might actually be those that suffer the most from unfair access to services. This phenomenon is known as the 'inverse care law' (Tudor-Hart 1971). Unfortunately, given the small number of diabetics registered with most of our surgeries, it is not possible to analyse the data in terms of equity regarding the different socio-economic groups or different ethnic groups. Therefore, we have only analysed the data in terms of equity of service utilisation across the 14 surgeries, for the whole set of diabetics. This allows us to evaluate the equity on a geographical basis and how it relates to the cost efficiency of the surgeries.

The trade-off between efficiency and equity has already been investigated by a number of other studies. Maital (1973, 1975) has suggested that an increase in the efficiency of public services provision tends to lead to a decrease in the equality of provision, benefiting those from higher socio-economic groups. Bodily (1978) used multiattribute utility theory to show this trade-off in terms of police services delivery. Mandell (1991) used the Gini coefficient to develop a model to examine the trade-offs between output maximisation and equality in the provision of public services. More recently, Golany and Tamir (1995) developed a DEA model that aims to allocate resources taking into account not just output maximisation considerations, but also equality and effectiveness considerations.

Our work differs from that of Golany and Tamir (1995) in terms of the definition of effectiveness. Whilst they define effectiveness in relation to the achievement of a certain level of output, we define it as the achievement of maximum outcomes from the outputs produced (services delivered). Furthermore, whilst they were concerned with equality of resource allocation between providers, we are concerned with equity of access to services between groups of patients. A fair allocation of resources between providers is a necessary, *but not sufficient* condition to guarantee equity of access to services across different groups of patients. Another aspect that distinguishes our work is the use of cost efficiency measures. These measures include both technical and allocative considerations. We feel that this is important in evaluating the potential trade-offs because achieving a higher level of equity may not only lead to an increase in the quantities of resources used, it may also lead to the use of more expensive resource mixes. Similarly, to achieve a higher level of effectiveness, a more expensive resource mix may be necessary.

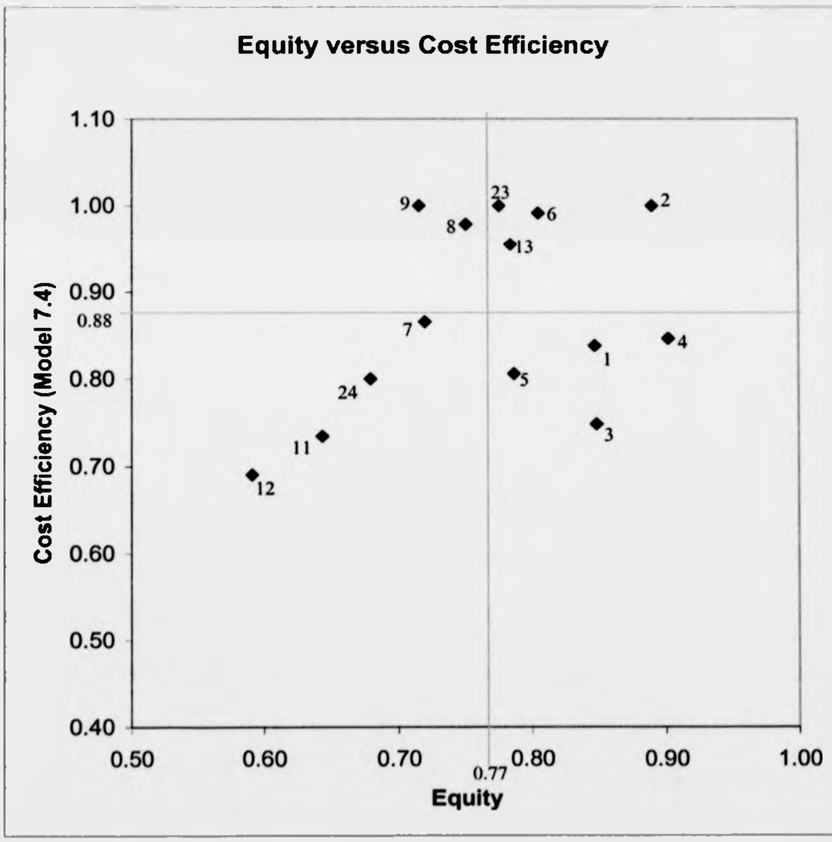
Figure 8.1 plots the results of cost efficiency against the results of equity for our 14 surgeries. This is a very important trade-off because, ideally, we would like to improve on both these performance criteria. In this respect, it is important to identify surgeries that are able to offer fair access to the essential services, whilst delivering the services in a cost efficient way.

If we take the sample as a whole, we cannot find a clear trade-off between these two performance measures. On the contrary, a weak positive relationship is noticeable. The Spearman rank correlation coefficient between the equity results and the cost efficiency results is 0.302, but it is not statistically significant at the 0.05 level (Table 8.1).

It is interesting to note that there is a group of surgeries for which a strong positive relationship between equity and cost efficiency is visible (surgeries 6, 7, 11, 12, 13 and

24). This may suggest that the relationship between these two measures might be contingent upon context. We obviously would need to further investigate this relationship in larger samples in order to be confident of this result.

Figure 8.1: Trade-off between equity and cost efficiency



If we take the average result of each surgery for each of the criteria and draw a line separating the surgeries that perform below average from those that perform above average, we can classify the surgeries into four groups.

The first group is composed by the surgeries that perform below average in both dimensions: surgeries 7, 11, 12 and 24. There is evidence that these surgeries could improve both in terms of equity of access to services and cost efficiency. They appear to deliver diabetes services to a relatively small proportion of their patients, whilst spending higher than average costs.

The second group is composed by the surgeries that perform above average in terms of equity of access, but below average in terms of cost efficiency: surgeries 1, 3, 4 and 5. These surgeries are able to regularly review a relatively high proportion of their diabetics, but appear to be doing this at a relatively high cost.

The third group is composed by the surgeries that perform above average in terms of cost efficiency and below average in terms of equity of access: surgeries 8 and 9. These surgeries appear to be cost efficient, but are not regularly reviewing a significant proportion of their registered diabetics.

The fourth group is composed by the surgeries that perform above average in both dimensions: surgeries 2, 6, 13 and 23. These surgeries review a relatively high proportion of their patients and are able to do this at relatively low costs. If we had not assessed performance from any other perspective, we would think that these are the surgeries that we need to learn from regarding diabetes service delivery. They show evidence of a sound system of diabetes delivery, which guarantees high accessibility of services at reasonable costs. However, we also need to evaluate how they score in terms of effectiveness. Do the systems used by these surgeries lead to good outcomes? For this analysis, we turn to the next section.

8.2.2 Trade-offs between Efficiency and Clinical Effectiveness

Schinnar et al. (1990) and Schinnar (1993) have pointed out the importance of investigating the trade-off between efficiency and effectiveness in public services delivery. As discussed in the literature review, in their study of mental health care programs, Schinnar et al. (1990) found a positive relationship between efficiency and effectiveness for lower than average levels of efficiency. Schinnar (1993) further investigates this issue and proposes that for different levels of efficiency the relationship may in fact change, with a positive association at low levels of efficiency and a negative association at higher than average levels of efficiency. In this respect, he suggests: "moderate efficiency levels may be indicative of higher effectiveness" (Schinnar 1993: 188).

We feel that it is of interest to analyse the relationship between cost efficiency and clinical effectiveness, because these two performance measures relate to different priorities. Whilst cost efficiency focuses on delivering the maximum amount of services at minimum cost, clinical effectiveness aims to improve the clinical outcomes derived from the services delivered. In many situations, the most efficient solution is not necessarily the one that will lead to the best outcomes. Again we have used cost efficiency rather than technical efficiency to account for the fact that input mix is of particular importance. An expensive mix might be what makes the difference in terms of outcomes achievement.

Figure 8.2 presents the relationship between the results obtained for cost efficiency and the results obtained for clinical effectiveness. From our results, we cannot identify a clear relationship between cost efficiency and clinical effectiveness. The Spearman rank correlation coefficient between our results for cost efficiency and our results for clinical effectiveness is – 0.205, but it is not statistically significant at the 0.05 level (Table 8.1).

Furthermore, our results do not confirm Schinnar's (1993) suggestion regarding different relationships at different levels of efficiency.

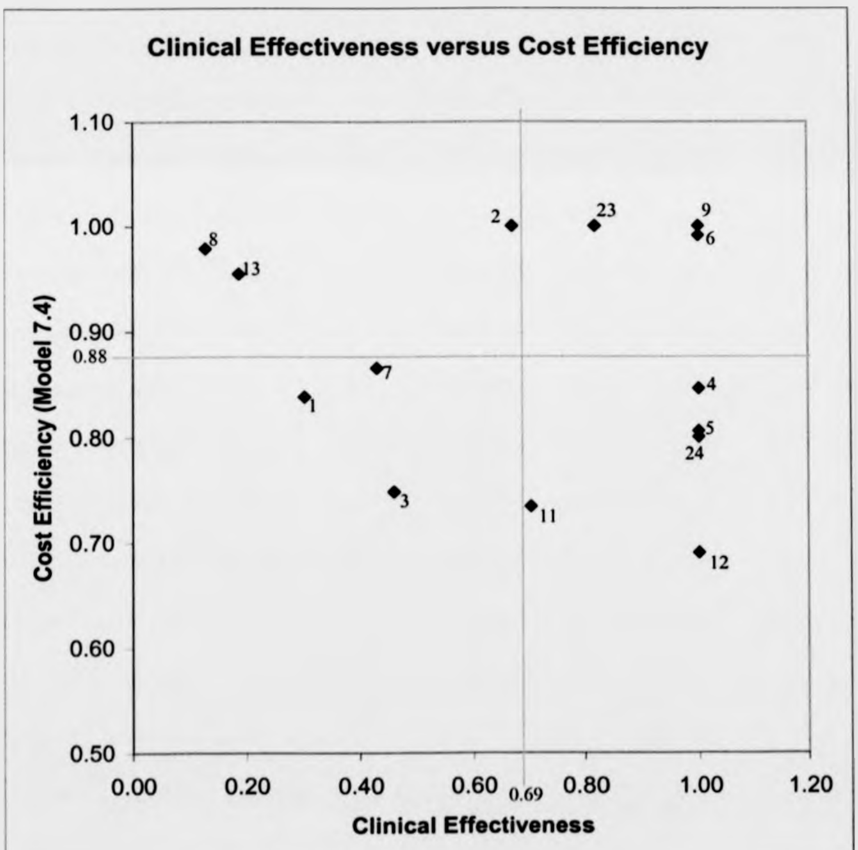
If we focus on Figure 8.2 we can once again identify four clear groups. The first group constituted by the surgeries that perform below average both in terms of effectiveness and efficiency: Surgeries 1, 3 and 7. This group of surgeries appear to show potential for improvement both in terms of cost efficiency and in terms of effectiveness. It is interesting to mention that all these surgeries score 100% in terms of technical efficiency. In this respect, they appear to be using a sub-optimal resource mix without actually achieving high levels of outcomes.

The second group is constituted by surgeries that perform below average in terms of efficiency, but above average in terms of clinical effectiveness: surgeries 4, 5, 11, 12 and 24. These surgeries appear to be doing 'the right things' in terms of achieving the clinical outcomes; however they appear to be operating with a very expensive model of care. Surgeries 5 and 12 suffer mostly from technical efficiency, which means that there is evidence that all the resources could be radially decreased whilst delivering the same quantities of services. On the contrary, Surgeries 4, 11 and 24 are suffering mostly from allocative inefficiency by using an expensive mix of resources.

The third group is constituted by surgeries that operate above average in terms of cost efficiency, but below average in terms of effectiveness: Surgeries 2, 8 and 13. These surgeries seem to be delivering services in the 'right way' in order to save resources, but they do not appear to be achieving good enough clinical outcomes from the services delivered.

Finally, we have a group of three surgeries that appear in the upper right quadrant, performing above average in both dimensions. These surgeries are surgeries 6, 9 and 23, which appear to be doing the 'right things in the right way'. That is, they are able to deliver services in a cost efficient way and at the same time, achieve a good level of clinical effectiveness. It is not surprising to find that these are the only three surgeries that score 100% in terms of cost effectiveness (Model 7.7).

Figure 8.2: Trade-off between effectiveness and efficiency



8.2.3 Trade-offs between Equity and Clinical Effectiveness

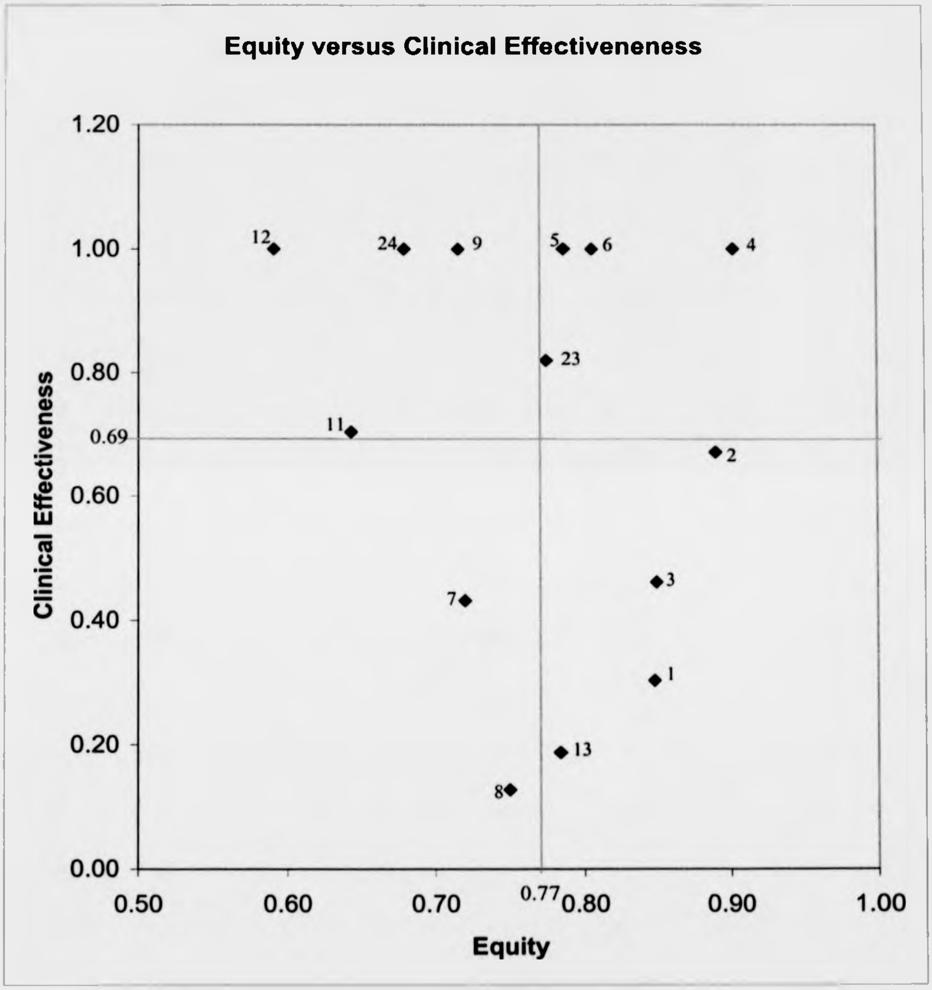
We feel that this potential trade-off between equity and clinical effectiveness is of particular significance for current health care policy in the UK, given the focus on outcomes achievement derived from the modernisation agenda of this Government. In terms of primary care delivery, we believe that it would be important to further investigate this trade-off in order to better understand the consequences that may result from the implementation of the new GP contract, which introduces financial rewards for high level of outcomes achievement (British Medical Association 2003).

During the workshops, a number of GPs emphasised the fact that one of the ways to increase effectiveness may be to reduce the proportion of patients to whom services are delivered on a regular basis. It seems that a potential unintended consequence of this might be another application of the 'inverse care law'. Those patients that most need health care assistance might find it more and more difficult to get appropriate care and dedication from the professionals. Smith (1993) also cautions to this potential unintended consequence of performance assessment systems that emphasise outcome measurement above everything else. We are not suggesting that all health care professionals would behave in this way, because personal gain is not the only value heuristic guiding their decisions. However, there might be some professionals that feel that it is too high a price to pay for following other values such as individual need and fairness.

The results of clinical effectiveness were plotted against the results of equity and the relationship between these two measures is presented in Figure 8.3. A clear trade-off between these two measures is not visible for the whole sample, which confirms that there may be other issues explaining the range of performance results obtained. In this respect, for the period of analysis considered, and for our sample of surgeries, a clear trade-off between equity and clinical effectiveness is not visible. The Spearman rank correlation coefficient between the equity results and the clinical effectiveness results is equal to – 0.11, but once again this result is not statistically significant at the 0.05 level (Table 8.1). It would be of interest to re-run this analysis at a later stage, after the full implementation of the new GP contract.

From Figure 8.3, we can identify four groups of surgeries depending on their quadrant location. The first group contains the surgeries that perform below average in terms of both criteria: surgeries 7 and 8. We then have a group of surgeries that score above average in terms of equity, but score below average in terms of clinical effectiveness: surgeries 1, 2, 3 and 13. These surgeries are regularly reviewing most of their diabetics, but they score below average in terms of the level of outcomes achievement. Surgeries 9, 11, 12 and 24 achieve a high level of clinical effectiveness but fail to review a significant proportion of their diabetics. Finally, we have a group of surgeries that are able to achieve high levels of clinical effectiveness, whilst reviewing most of their diabetics: surgeries 4, 5, 6 and 23. These surgeries can provide interesting models of diabetes care delivery. Surgery 4 is of particular interest because it achieves the highest score of equity, whilst scoring 100% in terms of clinical effectiveness. In the next chapter we shall try to identify the structures and mechanisms that are behind the successful performance of these surgeries.

Figure 8.3: Trade-off between equity and clinical effectiveness



8.2.4 Trade-off between Clinical and Patient Focused Effectiveness

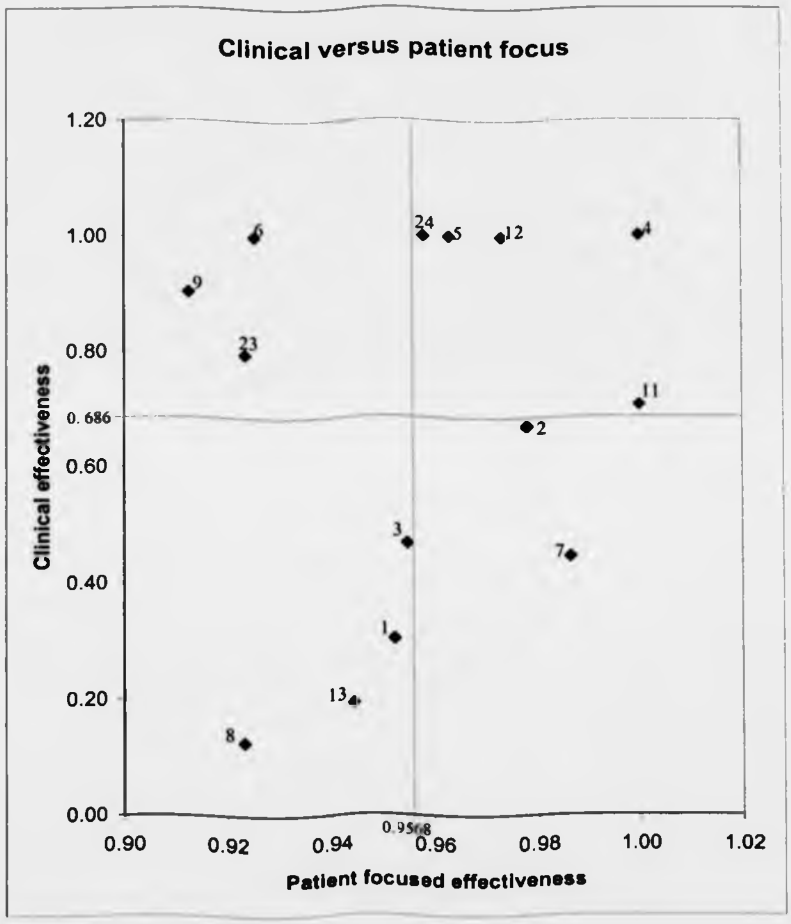
Finally, it is interesting to analyse the trade-off between scoring highly in terms of the achievement of clinical outcomes and scoring highly in terms of the achievement of patient satisfaction, understanding and confidence. The issue of patient satisfaction is of particular relevance here, because it is reasonable to expect that very intensive control of the clinical indicators might lead to lower levels of patient quality of life and satisfaction.

Figure 8.4 plots the results of clinical effectiveness against the results of patient-focused effectiveness. We can conclude that the results for these two measures are not concentrated around two quadrants, which means that a clear relationship between them is not applicable to the whole sample. The Spearman rank correlation coefficient between our results of clinical effectiveness and our results of patient-focused effectiveness is 0.239, but it is not statistically significant at the 0.05 level (Table 8.1).

Again we can find four distinct groups of surgeries depending on whether they perform below average in both dimensions (surgeries 1, 3, 8 and 13); whether they perform above average in terms of clinical effectiveness and below average in terms of patient-focused effectiveness (surgeries 6, 9 and 23); whether they perform above average in terms of patient-focused effectiveness but below average in terms of clinical effectiveness (surgeries 2 and 7); or whether they perform above average in both clinical and patient perspectives (surgeries 4, 5, 11, 12, and 24).

The interesting aspect to notice is that the surgeries that perform above average on both patient and clinical effectiveness; *all* perform below average in terms of cost efficiency. In fact, if we investigate the relationship between patient-focused effectiveness and cost efficiency, we find that these two variables are negatively correlated. The Spearman rank correlation coefficient is equal to – 0.539 and is statistically significant at the 0.05 level (Table 8.1). This is an interesting relationship, which suggests that there may be a trade-off between cost efficiency and patient-focused effectiveness. We would need to further investigate this aspect using a larger sample in order to be confident about this trade-off.

Figure 8.4: Trade-off between clinical and patient-focused effectiveness



8.2.5 Equity, Efficiency and Effectiveness – What is an acceptable balance?

After analysing the relationships between the different performance measures, we can identify some surgeries that need further investigation in order to explain their below average results in terms of performance. Firstly, we should further investigate Surgery 7, which scores below average in most criteria. Secondly, we should further investigate surgeries 1, 3, 8 and 13 as these surgeries score below average in terms of both clinical and patient effectiveness. In the next chapter we will be looking at some of the factors that may explain these results.

From another side, an analysis of the results from the different performance criteria reveals three surgeries that appear to be robust performers. Surgeries 6 and 23 score above average in all criteria apart from patient-focused effectiveness. If a compromise has to be struck between achieving a high level of equity, delivering services in a cost efficient way, and achieving a good level of effectiveness, these surgeries appear to be good learning sources. However, if we 'ignore' costs for a moment, then Surgery 4 appears to offer a very good example of diabetes care delivery. This surgery has the highest score in terms of equity (above 90% of the patients receive all the diabetes services on time) and it scores 100% in terms of both clinical and patient-focused effectiveness. Furthermore, when we look at pure technical efficiency it also scores 100%. The only aspect that penalises its performance is that it appears to be using a relatively expensive input mix in order to achieve these good results. The average estimated cost per diabetic in Surgery 4 is £711 per year, whilst in Surgery 6 this cost is reduced to £580. However, analyses of the results from the patient questionnaire reveal that this saving has a price in terms of reduced patient responsiveness. In Surgery 6, 33% of the patients report unacceptably long waiting times to see a doctor and 28 % of patients say that they do not have enough time to talk with the doctor.

The question is how much cost efficiency are we willing to sacrifice in order to achieve the highest level of patient satisfaction, fairness and clinical effectiveness? This is certainly a political question and a full discussion of this topic is outside the scope of this thesis. However, we feel that it would be important to further investigate these trade-offs in order to make informed policy choices regarding resource allocation and primary health care delivery.

8.3 Robustness of the Results – The Addition of nine Surgeries to the Sample

One of the limitations of this study relates to the small size of the sample under analysis. It has been suggested in the literature that the robustness of DEA results is a function of the size of the sample under analysis. In general, smaller samples tend to be less robust (Pedraja-Chaparro, Salinas-Jiménez and Smith 1999). In this section we have introduced data for 9 additional surgeries and have analysed the robustness of our performance results. Please refer to Appendix G for the data referent to these additional 9 surgeries.

As previously discussed, we collected data for 23 surgeries and it was our initial objective to analyse the data for all these surgeries. However, due to concerns about the quality of the data for some of these surgeries and about the homogeneity of the sample as a whole, we chose to concentrate on the analyses of 14 surgeries.

It is important to note that the purpose of this study is formative evaluation. It is not our intention to obtain a ranking of the surgeries in the sample, nor it is our intention to obtain a good estimator of the average efficiency or effectiveness in primary diabetes care delivery. We are interested in using the DEA results to learn about better ways to deliver diabetes care. We aim to identify surgeries that appear to be under performing in a particular dimension and make suggestions on how their performance could be improved. This means that we are interested in identifying surgeries that are all round good performers and that often act as peers for other surgeries.

As previously discussed, Surgeries 4, 6 and 23 appear to be 'robust performs', because they score above average in most performance criteria and are often the learning peers to other surgeries. We will now discuss whether the inclusion of other surgeries has penalised these 'robust performers'. We emphasise that we are not particularly interested in the individual scores of the surgeries. However, we are interested in the changes in the patterns of the results, as whole. In this respect, and in order to maintain consistency with the previous section, we have chosen to present the results using trade-off graphs. Figures 8.5, 8.6 and 8.7 present the results for the 23 surgeries.

Analyses of these three graphs reveal that our performance results are generally very robust. Most of our 14 surgeries maintain their quadrant location. The only exception is for surgeries whose scores were initially very close to the average, which have seen their scores change slightly. In this case, a few surgeries have changed from one quadrant to the adjacent quadrant, but still remain close to the average line. Surgery 23's equity score drops from slightly above average to slightly below average. With that exception alone, our 'all round good performers' maintain their status and appear predominantly in the upper quadrant. Surgery 4 continues to score very highly in terms of equity and effectiveness. If we analyse the results from the perspective of relatively poor performers, such as Surgery 7, we find that this surgery continues to appear in the lower left quadrant of the three graphs.

Finally, it is of interest to analyse the number of times that a surgery appears as peer to others, when we consider our objective of identifying 'role model' surgeries. Both Surgery 6 and Surgery 23 continue to be frequent peers to other surgeries in terms of cost efficiency and cost effectiveness. Surgery 12 continues to be the most frequently chosen peer in terms of clinical effectiveness. It is now a peer to 19 surgeries. Surgery 4 remains effective from a patient perspective and is peer to 12 surgeries. However, this surgery is no longer part of the frontier in terms of clinical effectiveness. Nevertheless, it still scores very highly in this dimension.

When we consider our concerns regarding the quality of the data for these additional 9 surgeries, we shall not further analyse the results for the extended sample. Our intention is to concentrate on the results from our sample of 14 surgeries in order to gain some insights into diabetes service delivery and primary health care policy. This comparison with the results from the extended sample has reassured us regarding the relative robustness of the initial results. Furthermore, we would not suggest the use of these additional 9 surgeries as examples of 'best practice' due to the lack of confidence in their data. In the next chapter we will investigate the impact of some environmental variables on these performance measures.

Figure 8.5: Trade-off between equity and cost efficiency (23 surgeries)

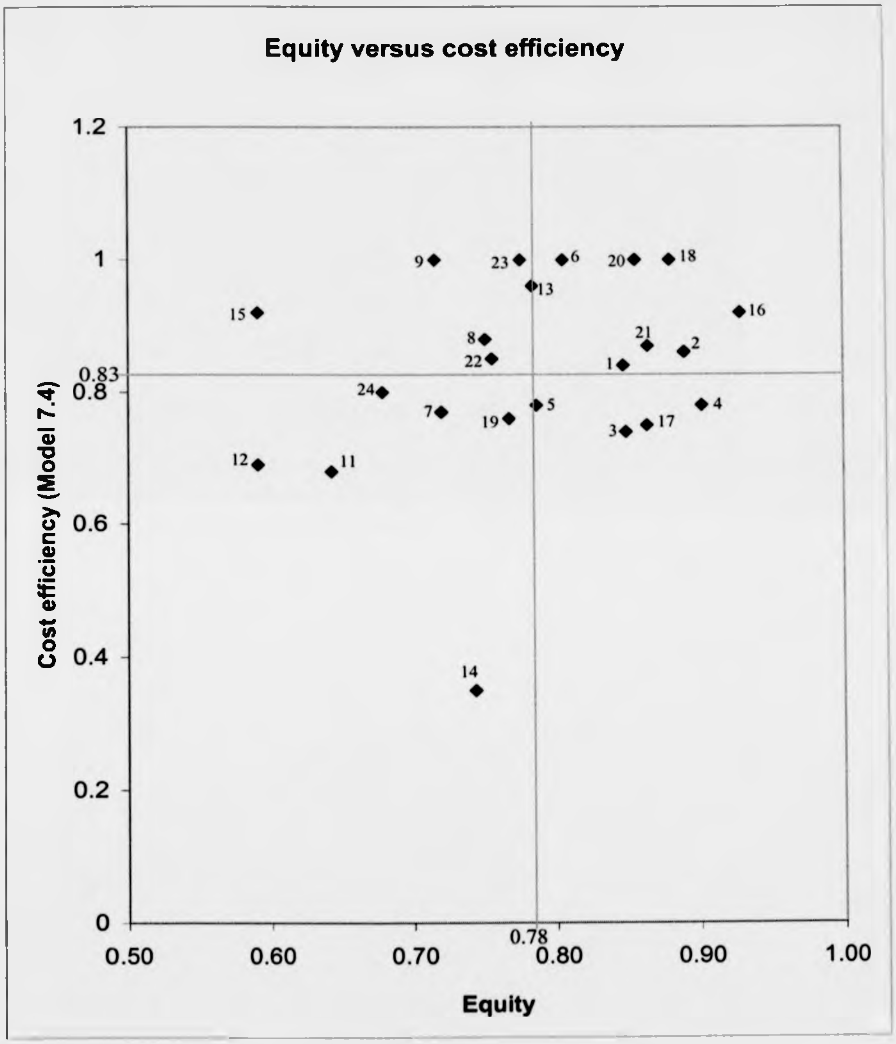


Figure 8.6: Trade-off between clinical effectiveness and cost efficiency (23 surgeries)

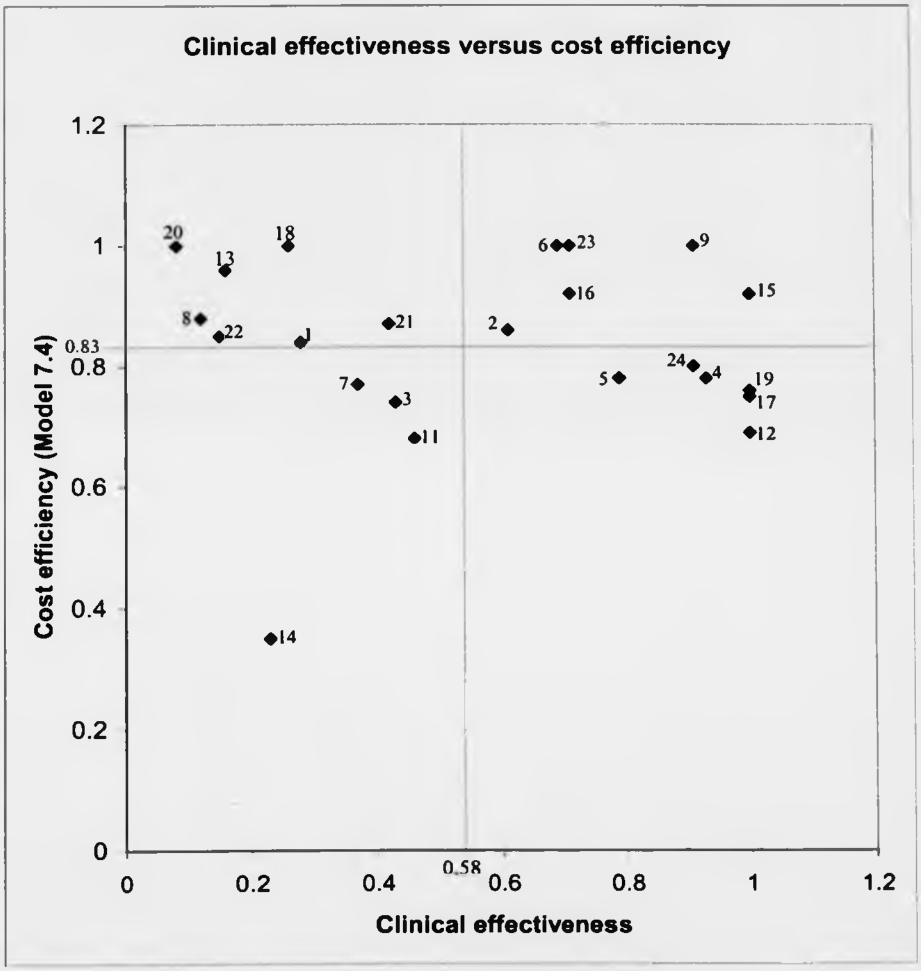
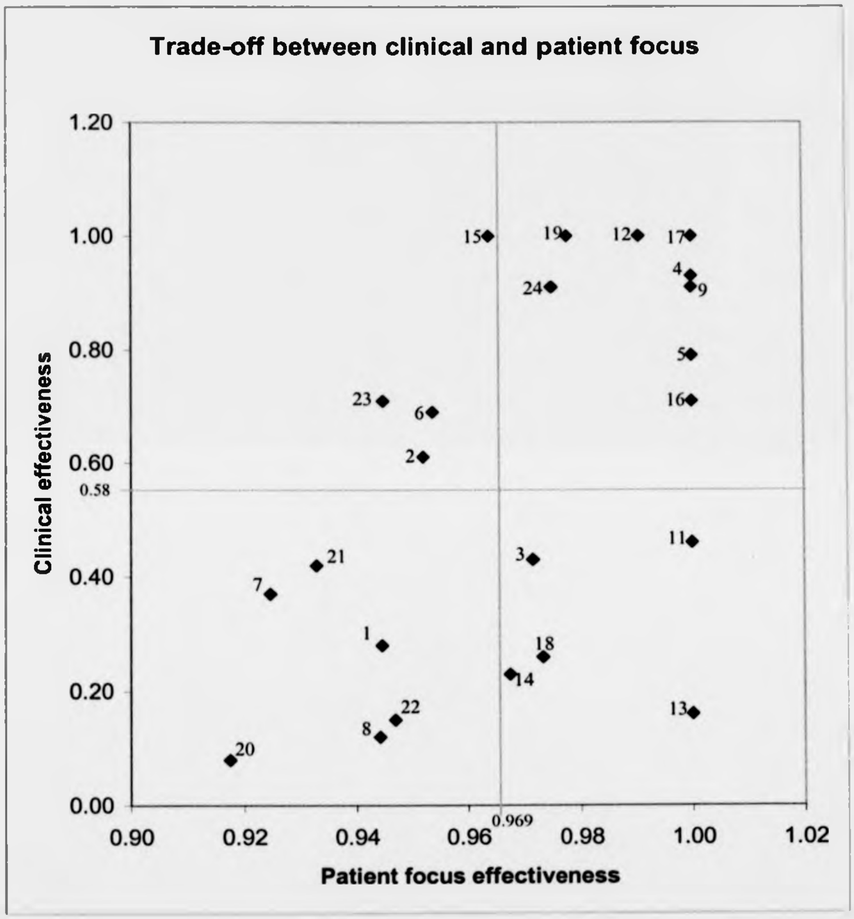


Figure 8.7: Trade-off between the clinical and patient-focused effectiveness (23 surgeries)



8.4 Conclusion

In this chapter we have discussed the results from the diabetes case study. We have analysed a range of alternative performance measures, as well as the trade-offs between them. For our sample of 14 surgeries, we could not identify significant relationships between equity, cost efficiency and clinical effectiveness. A significant negative Spearman rank correlation was found between cost efficiency and patient-focused effectiveness.

Based on our results for the different performance criteria, we conclude that some surgeries appear to be able to achieve a good balance between the major performance criteria considered. Surgery 4 scores above average in all performance criteria apart from cost efficiency. Surgeries 6 and 23 score above average in all performance criteria apart from patient-focused effectiveness. These three surgeries appear therefore to provide examples of good practice in diabetes care delivery and management.

From another perspective, there are surgeries that score highly in one dimension but score poorly on a different dimension. Surgery 12, for example, is 100% clinically effective, but presents the lowest score in terms of cost efficiency and equity. Additionally, there is a group of surgeries that appear to show potential for improvement along most of the performance dimensions. For example, surgery 7 presents below average results for all performance criteria apart from patient-focused effectiveness. What remains to be explained is whether some of the surgeries operate in fundamentally different environments, and therefore should not be compared with the others, or whether they do not have adequate systems and mechanisms in place in order to deliver good diabetes care. We will try to answer these questions in the next chapter.

Before we move onto this analysis, it is worth emphasising that we cannot ignore the complexities involved in performance assessment and the difficulties in deciding *which* *surgeries represent* 'best practice', and *which surgeries need* to change their way of working. Once we recognise that decision-making in the public sector is guided by several distinct (and potentially conflicting) values, we also need to recognise that performance assessment cannot lead to a 'black and white' classification.

9 Post Evaluation: What can we Learn from the Results?

9.1 Introduction

This chapter critically analyses the results and discusses their validity and usefulness in practice. In the first part of the chapter we return to the trade-off analyses graphs regarding the different performance assessment criteria and bring in some of the environmental variables, in order to discuss their impact on the results. We analyse two types of environmental variables: area related factors and patient related factors. Furthermore, we analyse the effect of different structures and processes of diabetes care delivery on the performance results. The fact that we compared a small sample of surgeries did not allow the inclusion of these factors into the DEA models.

In the second part of the chapter we discuss the results in light of the workshops undertaken with some of the surgeries under evaluation, as well as the workshops with South Peterborough PCT. The aims of these workshops were the following. Firstly, we wanted to identify the 'reputational perspective' regarding performance. Secondly, we wanted to 'validate' the results by discussing them and their meaning with the surgeries. Thirdly, we wanted to identify some of the structures and mechanisms behind the performance results. Finally, we wanted to develop plans of action to improve the performance of the participating surgeries, and identify ways in which the results of this study could be used to develop strategies for a more equitable, efficient and effective diabetes care delivery.

9.2 The Impact of Environmental Variables on Performance

We have analysed the impact of two types of environmental variables: (1) area related factors and (2) patient related factors. Some of these variables are categorical, whilst others are numerical. In order to evaluate the impact of categorical variables we have clustered the surgeries into sub-sets regarding their characteristics and then have looked for patterns in the performance results within and between the sub-sets. In order to evaluate the impact of numerical factors, we have calculated the Spearman rank correlation between each of the variables and the results for the different performance criteria. In this case, the Spearman rank correlation coefficient was chosen because the performance scores are not normally distributed. Therefore, it is preferable to calculate the correlation between the ranks of the measures than between the measures themselves.

9.2.1 Area Related Factors

In terms of area related factors, we used the level of area deprivation to distinguish between surgeries facing a high, medium and low level of area deprivation. We have also grouped the surgeries according to their location: rural, suburban, inner city and mixed.

9.2.1.1 Area Deprivation

Area deprivation is a very important factor in primary health care delivery, when we consider that it accounts for the patients' relative socio-economic disadvantage. Several UK studies have suggested that morbidity and mortality levels associated with diabetes tend to be worsened by socio-economic deprivation (Roper et al. 2001; Acheson Report

1998; Urwin et al. 1996; Connoly and Kesson 1996 and Kelly et al. 1993). In this respect, we have decided to analyse the impact of area deprivation on the classification of the surgeries.

The measure used to represent socio-economic deprivation was derived using the methodology developed by the Department of the Environment, Transport and the Regions (DETR 2000), regarding their research in terms of Indices of Multiple Deprivation (IMD) for wards. "The IMD 2000 score is the combined sum of the weighted, exponentially transformed domain rank of the domain score" (DETR 2000: 14). Six domains are used with the following weights: Income (25%); Employment (25%); Health deprivation and disability (15%); Education skills and training (15%); Housing (10%) and Geographical access to essential services (10%).

In order to obtain a surgery score, the patients registered with each surgery were assigned to a ward based on their postcodes, and an average IMD score was then calculated for each surgery using the ward scores obtained by DETR (2000). The higher the IMD score, the higher the level of deprivation. However, given that the multiple score results from a combination of measures which have been transformed, this score is not a ratio variable and we cannot say for example that a surgery with a score of 30 is twice as deprived as a surgery with a score of 15. For this reason, and for clarity of analysis, we opted to classify the surgeries into three groups regarding their level of deprivation: low, medium and high. We first calculated the Median IMD score for the surgeries under evaluation and then classified surgeries with an IMD score below the median IMD score as facing low deprivation. Surgeries at the median IMD as facing a medium level of deprivation and surgeries with an IMD score above the median as facing a high level of deprivation. Unfortunately, an IMD score could not be obtained for surgeries 23 and 24 because the data was not available. The classification of these two surgeries is therefore based on the judgement of the health care professionals working at each one of these surgeries. Table 9.1 presents the classification in terms of area deprivation. This classification was discussed during the workshops and it was felt to group correctly the surgeries into different levels of deprivation.

We have used this classification to make new maps regarding the trade-offs between the several performance measures in order to evaluate the patterns that occur. Three trade-off graphs are presented below (Figures 9.1 to 9.3).

Table 9.1: Level of deprivation

Surgery	Level of area Deprivation
1	High
2	Low
3	High
4	Low
5	Medium
6	Medium
7	Low
8	High
9	Medium
11	Low
12	Medium
13	High
23	Medium*
24	Medium*

* Estimated classification by the health care professionals of the surgery

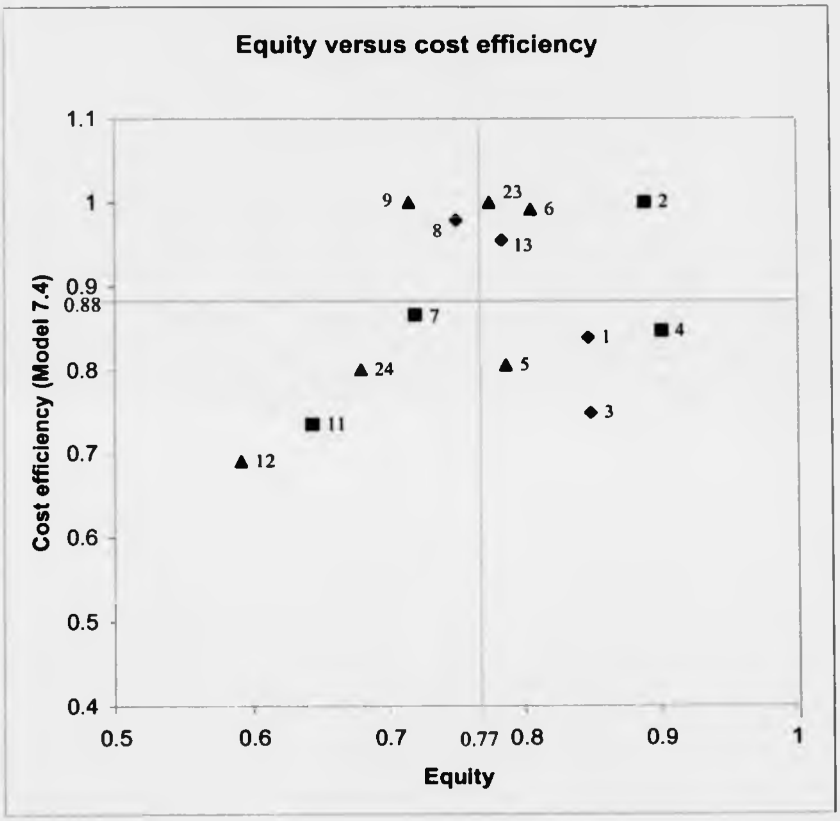
The analyses of these three graphs suggest that a high level of area deprivation may explain the low scores of clinical, patient-focused and cost effectiveness obtained by certain surgeries. We can conclude that *all* the surgeries facing high levels of area deprivation appear to perform below average in terms of clinical, patient-focused and cost effectiveness. The Acheson Report (1998) suggested that poor BP control can be a proxy for high population deprivation. We can also note that *all* surgeries located in areas of low deprivation present a higher than average level of patient-focused effectiveness. Given that two of the outcomes included in this model were the level of patient understanding and the level of patient confidence, this result is not surprising.

Surgery 8 is a good example of how difficult it may be to deliver services for a highly deprived population. This relatively small surgery (two GPs and one practice nurse), covering a highly deprived area, presents the lowest score of clinical effectiveness. Further analyses, based on the results from the patient questionnaire, also reveal that this surgery faces the highest proportion of unemployed patients and the highest proportion of long-term sick and disabled patients in the sample. Despite the fact that this relationship between high deprivation and poor effectiveness is consistent with the results from previous studies (Roper et al. 2001; Urwin, et al. 1996; Connoly and Kesson 1996 and Kelly et al. 1993), further investigations in larger samples would be required in order to better investigate this relationship.

Another issue of interest in terms of the effects of deprivation in health care provision relates to the level of services utilisation between different socio-economic groups. Le Grand (1978) suggested that patients from higher socio-economic groups tended to have higher rates of service utilisation than those from lower socio-economic groups. Our results in terms of equity do not appear to confirm this suggestion. Two of our surgeries located in low deprivation areas present a lower than average level of equity and three of the surgeries covering highly deprived areas present a higher than average level of equity.

However, it is interesting to notice that based on our estimates, the average annual cost of diabetes varies between surgeries serving low deprivation areas (£717.14 per patient); medium deprivation areas (£626.60 per patient) and surgeries serving highly deprived areas (£600.49 per patient). One comment made by one specialist diabetes nurse might shed some light into this. She suggested that highly informed, highly educated patients tend to *demand* more expensive drugs and more frequent consultations in order to monitor their condition. For our sample, the average cost of prescriptions issued to each diabetic patient in surgeries covering areas of low deprivation is £363.46, whilst it only amounts to £287.41 in surgeries covering highly deprived areas. However, due to the small size of our sample, it is not possible to make firm conclusions regarding the meaning of these results. It would be interesting to further investigate this issue.

Figure 9.1: The impact of deprivation on equity and cost efficiency

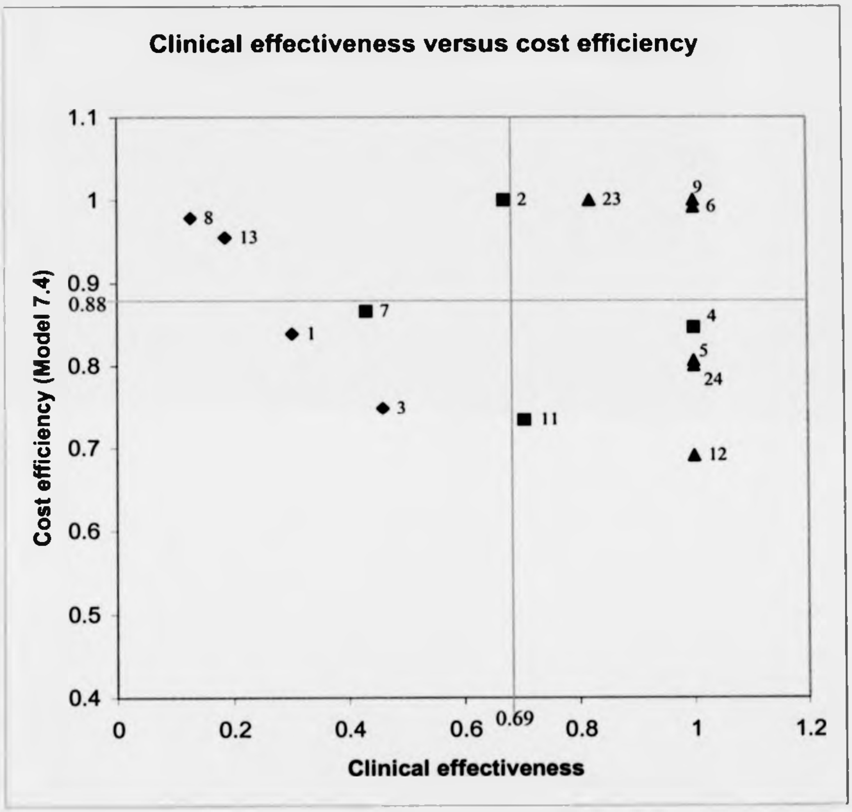


Key ■ Area of low deprivation

▲ Area of medium deprivation

◆ Area of high deprivation

Figure 9.2: The impact of deprivation on effectiveness and efficiency

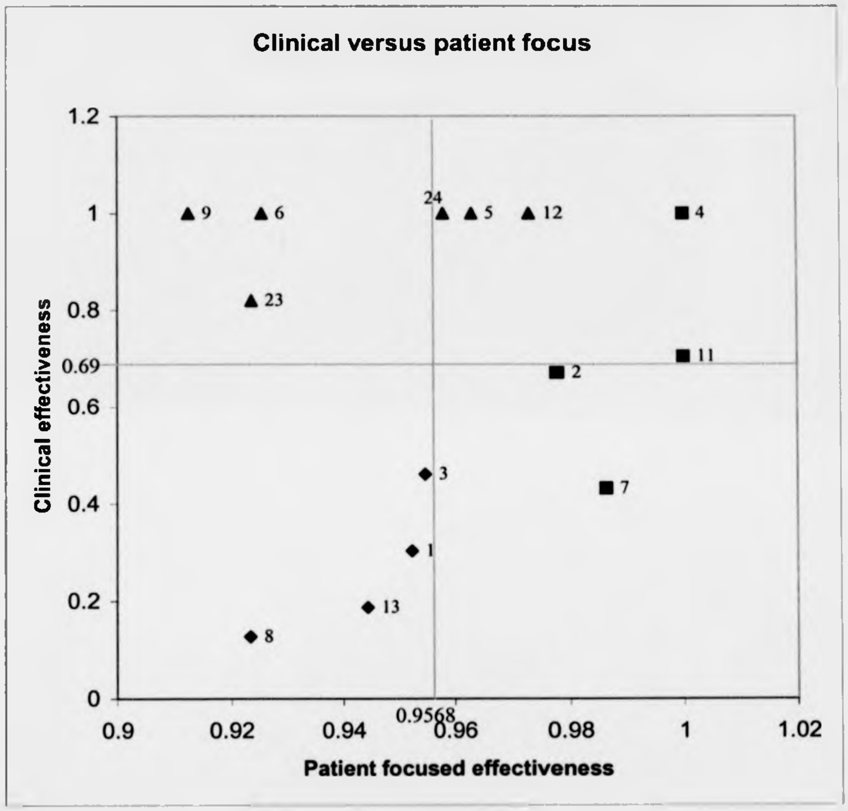


Key ■ Area of low deprivation

▲ Area of medium deprivation

◆ Area of high deprivation

Figure 9.3: The impact of deprivation on clinical and patient-focused effectiveness



Key ■ Area of low deprivation

▲ Area of medium deprivation

◆ Area of high deprivation

9.2.1.2 Location

We would expect the location of a surgery to have an impact on the way it operates. It is plausible to think that surgeries located in a rural area will have to operate rather differently than those located in an inner-city area. Furthermore, patients' area of residence can have an effect on healthcare utilisation and health outcomes (Mitchell et al. 2000, Gilthorpe and Wilson 2003). From our 14 surgeries, two surgeries are located in an inner-city location; three are located in a suburban location; four are located in a mixed location and the remaining five are located in a rural area. Table 9.4 presents the location of each surgery.

Table 9.4: Location

Surgery	Location
1	Mixed
2	Rural
3	Suburban
4	Rural
5	Rural
6	Mixed
7	Rural
8	Inner city
9	Mixed
11	Rural
12	Suburban
13	Inner city
23	Suburban
24	Mixed

In order to analyse if location has an impact on the performance measures obtained, we have once again re-mapped the surgeries in terms of their performance scores, making their location explicit. These analyses do not suggest any relationship between location and our measures of performance, and due to space limitations, we have opted to leave these maps out. The only pattern that we can note is that all surgeries located in rural areas score above average in terms of patient-focused effectiveness. This means that, on average, these surgeries score better in terms of patient understanding, confidence or satisfaction. Given that most of our surgeries located in rural areas face a low level of deprivation, this may just say something about the type of population covered by these surgeries and their motivation in terms of diabetes control. In order to investigate the impact of patients' characteristics on the performance results, let us now turn to patient related factors.

9.2.2 Patient Related Factors

We have attempted to analyse the impact of the patient related factors by looking at: (1) the level of patient compliance, patient confidence and patient understanding of diabetes; (2) the proportion of elderly patients; the proportion of newly diagnosed patients; the proportion of patients with type-1 diabetes and the proportion of patients with diabetes related complications.

9.2.2.1 The Effect of 'Patient Cooperation' on Performance

The level of patient compliance, confidence and understanding are three important factors in chronic disease management. These are both inputs and outcomes of diabetes care. From one perspective, they are inputs because a greater level of patient compliance, confidence and understanding can help in the achievement of better clinical outcomes. However, from another perspective, the quality of the diabetes care in itself will influence the level of patient compliance, confidence and understanding of diabetes. We have used these three variables as proxies for the level of patient cooperation in diabetes management. Table 9.5 presents the data for these three variables and Table 9.6 presents the Spearman rank correlations between these variables and the performance results[21]. Based on the correlation results, we can conclude that the only variable that presents a significant correlation with the performance results is the level of patient understanding, which is positively correlated with patient-focused effectiveness. This result is not surprising because patient understanding was one of the outputs used to measure patient-focused effectiveness. We will now discuss each one of the hypothesised relationships in more detail.

Table 9.5: Patient cooperation in diabetes management

Surgery	% of smokers who understand the effects of smoking	% of confident and very confident patients	% of patients who say they understand enough about diabetes
1	35.00%	87.60%	77.82%
2	39.00%	90.60%	80.98%
3	25.00%	94.19%	80.66%
4	26.00%	94.70%	80.52%
5	20.00%	92.00%	75.60%
6	35.00%	89.73%	78.24%
7	26.00%	88.00%	82.80%
8	29.00%	88.46%	76.98%
9	19.00%	89.72%	74.32%
11	17.00%	100.00%	84.55%
12	57.00%	81.82%	76.63%
13	27.00%	94.44%	76.90%
23	38.00%	86.55%	76.00%
24	16.00%	90.23%	77.44%
Average	**29.21%**	**90.57%**	**78.53%**

[21] The Spearman rank correlation coefficient calculates the correlation between the ranks of the different measures. This coefficient was used because the performance results follow a non-normal distribution.

Table 9.6: Spearman's rank correlations between the different performance criteria and patient cooperation

	Equity	Cost efficiency	Clinical effectiveness	Patient focused effectiveness
Patient non-compliance	0.249 (0.391)	0.316 (0.270)	-0.149 (0.612)	-0.136 (0.644)
Patient confidence	0.323 (0.260)	-0.205 (0.481)	0.052 (0.861)	0.409 (0.146)
Patient understanding	0.244 (0.401)	-0.223 (0.444)	-0.241 (0.408)	0.645* (0.013)

P-Values[22] given in parenthesis

* Correlation is significant at the 0.05 level (2 tailed)

There are two types of non-compliance: (1) voluntary (results from patient choice) and (2) involuntary (results from mistakes or lapses in memory). Both types of compliance are important factors that may influence clinical and cost effectiveness. However it is very difficult to assess any of these types. We have attempted to assess the level of voluntary non-compliance, using the data from the patient survey. The patients were asked two separate questions and the answers to these questions were collated to infer about the level of voluntary non-compliance. We have used the proportion of smokers who say that they understand the impact of smoking on their diabetes as a proxy for voluntary non-compliance. It is our conviction that surgeries with a higher proportion of non-compliant patients face more difficulties in achieving good clinical outcomes.

A similar argument could be made regarding the level of patient confidence. Patients who do not feel confident in taking a good control of their diabetes may lack the

[22] Probability that the observed relationship between the variables occurred by chance.

motivation to follow the diet and medication. Therefore, surgeries with a larger than average proportion of patients with low confidence may find it more difficult to deliver effective care.

In our small sample, it is not possible to identify any clear evidence regarding the influence of patient non-compliance and confidence on the performance results. Please refer to Table 9.6 for the Spearman correlations between these variables and the performance results for the different criteria.

The level of patient understanding is another factor that could explain the willingness of the patients to follow the professionals' instructions, and could therefore lead to a higher level of services utilisation and better clinical outcomes. Surgeries 7 and 11 have the highest proportion of patients reporting a good understanding of diabetes. These are two surgeries located in a rural area, with a low level of deprivation. However, neither one of them appears to distinguish itself in terms of good performance. On the contrary, Surgery 9 presents the lowest level of patient perceived understanding and still manages to be cost effective. When we analyse the correlations between patient understanding and the results for the different performance criteria, we conclude that they are not statistically significant (Table 9.6).

9.2.2.2 The Effect of Patient Characteristics on the Performance Results

The characteristics of the patients registered with each surgery can have an impact on the level of spending, services delivered and outcomes achieved. The inclusion of measures related to the case-mix of the patients registered with each surgery would have led to a fairer comparison between the surgeries. Unfortunately, the fact that we have compared a small sample of surgeries placed limitations on the number of variables that we could include in the DEA models. In this respect, in order to evaluate the impact of patient case-mix, we have investigated the correlations between several patient related factors and the performance results. In terms of diabetes care, we believe that there are five patient related factors that may impact on the performance of the surgeries: (1) the proportion of patients from ethnic minorities; (2) the proportion of elderly patients; (3) the proportion of newly diagnosed patients; (4) the proportion of type-1 patients and (5) the proportion of patients with complications. Table 9.7 presents the data for these variables and Table 9.8 presents the Spearman rank correlations between each one of these variables and the performance results[23]. Based on the correlation results, we can conclude that, for our sample of surgeries, none of these variables is significantly correlated with the results for the different performance criteria evaluated.

Table 9.7: Patient characteristics

Surgery	% of elderly patients (aged 75 years old or more)	% of newly diagnosed patients	% of patients with type-1 diabetes	% of patients with diabetes related complications
1	25.12%	11.27%	9.16%	48.12%
2	21.39%	14.16%	8.48%	53.76%
3	15.64%	18.01%	4.65%	46.45%
4	25.52%	18.88%	3.27%	58.04%
5	23.83%	15.42%	1.94%	54.21%
6	18.00%	18.44%	8.05%	63.77%
7	24.14%	17.24%	8.00%	48.28%
8	26.32%	15.79%	7.69%	45.61%
9	24.69%	9.62%	4.59%	54.81%
11	20.00%	28.00%	0.00%	64.00%
12	11.76%	17.65%	13.64%	33.33%
13	22.73%	38.64%	5.40%	46.59%
23	19.11%	12.53%	2.50%	59.24%
24	27.06%	10.19%	5.96%	53.70%
Average	**21.81%**	**17.56%**	**5.95%**	**52.14%**

[23] The Spearman rank correlation coefficient calculates the correlation between the ranks of the different measures. This coefficient was used because the performance results follow a non-normal distribution.

Table 9.8: Spearman's rank correlations between the different performance measures and patient characteristics

	Equity	Cost efficiency	Clinical effectiveness	Patient focused effectiveness
Proportion of elderly patients	0.024 (0.935)	0.148 (0.614)	-0.146 (0.618)	-0.066 (0.823)
Proportion of newly diagnosed patients	0.130 (0.659)	-0.316 (0.272)	-0.088 (0.766)	0.396 (0.161)
Proportion of type-1 patients	0.037 (0.899)	0.038 (0.899)	-0.178 (0.544)	-0.066 (0.823)
Proportion of patients with complications	0.090 (0.759)	0.298 (0.301)	0.452 (0.105)	0.132 (0.653)

P-Values[24] given in parenthesis

Firstly, the proportion of patients from ethnic minorities is one of the factors that may partly explain the performance results in diabetes care, when we consider the difficulties that the patients may have in understanding the language and when we consider their different cultures and lifestyles. The proportion of patients from ethnic minorities in all the surgeries compared in this study is very small (less than 1%) and, for this reason, we believe that this factor does not significantly affect the performance results.

Secondly, the proportion of elderly patients (patients aged 75 years old or more) can have an impact on the level of resource usage and the achievement of outcomes. The management of diabetes in elderly patients tends to be more difficult. A number of health complications may be present, such as poor eyesight and poor mobility and these can place additional difficulties in metabolic control. More frequent consultations may

[24] Probability that the observed relationship between the variables occurred by chance.

therefore be necessary in order to adequately manage their diabetes. We have calculated the Spearman rank correlation coefficient between the proportion of elderly patients registered with each surgery and the results for each of the performance criteria and cannot find any significant correlation between them (Table 9.8).

Thirdly, the proportion of newly diagnosed patients (patients diagnosed less than one year ago) is another factor that can explain the performance results. Newly diagnosed patients represent a different group of patients. They require longer and more frequent consultations until they adjust to the condition. Furthermore, it is possible that their metabolic values have not yet had a chance to respond to medication. In our sample, the proportion of newly diagnosed patients registered with each surgery varies considerably. Surgery 9 presents the lowest proportion of newly diagnosed patients (9.62%), whilst surgery 13 presents the highest proportion (38.64%). However, when we analyse the results for the sample as a whole, we cannot identify any significant correlation between this variable and the performance results (Table 9.8).

Fourthly, we have investigated the impact of the proportion of patients with type-1 diabetes[25] on the performance results. Type-1 patients require treatment with insulin, which is more expensive. Furthermore, the majority of type-1 patients are seen by a hospital diabetologist, which may result in a different level of service delivery and outcomes achievement. For these reasons, it is of interest to investigate whether surgeries with a larger proportion of type-1 patients perform significantly different on a particular criterion. We have analysed the data for this variable and we cannot identify any significant correlation between this variable and the performance scores (Table 9.8).

---

[25] We have used the data from the patient questionnaire to estimate the proportion of type-1 diabetic patients for each surgery. This was estimated by calculating the proportion of patients aged 40 years old or less, who are treated with insulin.

Finally, in order to account for the quality of the diabetes care provided in the past, we have estimated the proportion of patients who have developed diabetes related complications. A high proportion of patients with diabetes related complications may suggest a relatively difficult mix of patients, and may also lead to the practise of reviewing them more often, impacting on cost efficiency. Once again, it is not possible to find a correlation between this variable and the performance results (Table 9.8). For example, surgery 6 has the highest proportion of patients with complications and is nevertheless one of the best performers. On the other hand, surgery 12 has the lowest proportion of patients with complications and is inefficient (this surgery has the highest WTE GP per patient).

The difficulty in estimating the effects of environmental variables on the performance of primary care providers may be explained by the fact that in most cases the effect results from the combination of the several variables. Different factors interact to impact on the resources usage, service delivery and outcomes achieved and their effect is only visible if a multi-level system of equations is used (Salinas-Jiménez and Smith 1996). Johnson et al. (2002) in their study of variation of diabetes outcomes concluded that significant variability in glucose control was still found after it was taken into account for different levels of severity, compliance and knowledge, suggesting that other factors may influence the achievement of diabetes outcomes.

9.3 Surgery Related Factors: The Impact of Structure and Processes of Care

We can expect that the performance results obtained be partly explained by the different structures and processes in use in each surgery. We now return to the characteristics of these 14 surgeries and investigate the apparent relationship between different structures and processes, and the results in terms of the different performance criteria evaluated.

In terms of surgery related factors, we have grouped the surgeries depending on whether they appear to operate the diabetes services based on a GP-led, a nurse-led or a balanced system. We have also investigated which surgeries have a professional with formal training in diabetes, and whether this is the professional who delivers most of the diabetes services. Furthermore, we investigate whether a particular professional has ownership and autonomy in terms of diabetes care by comparing the professional who runs the diabetes clinics with the professional identified as the formal lead for diabetes.

We also grouped the surgeries into training and non-training surgeries, and into surgeries working under Personal Medical Services (PMS) contracts or working under General Medical Services (GMS) contracts, and investigated the impact of these two factors on the performance results.

9.3.1 GP-led versus Nurse-led Services

One aspect that can have an impact on performance is the success of the teamwork between the health care professionals operating at each surgery. Of particular relevance to primary care is the partnership between GPs and nurses, and the way both types of professionals work in order to better meet the needs of the patients at the lowest possible cost. Ovhed et al. (2000) investigated the importance of the role of the diabetes nurse in primary diabetes care in Sweden and concluded that primary care teams where nurses had a leading role performed better than those where nurses only had an assistant role.

In our sample of surgeries we have come across three different models of operation regarding the role of the nurses in diabetes care delivery. Whilst in some of the surgeries, the nurses have a leading role, running specialised diabetes clinics, prescribing medication and educating the patients; in other surgeries they only have a supplementary role of assistance to the GPs. In others surgeries a more 'balanced model' is used, where both GPs and nurses are involved in diabetes care.

We used the responses from the Practice Profile Questionnaire (Appendix E), in order to characterise the different models of diabetes care delivery. Firstly, we asked each surgery to say if they have a leading person for diabetes care, and if the answer was yes, we asked for the name and profession of this person. We defined this individual as the 'formal lead person' for diabetes care. Secondly, we asked each surgery to say if they run specific diabetes clinics, and whether they are GP-led, nurse-led or both. Thirdly, we asked the surgeries to describe the level of formal training in diabetes of the GPs and nurses involved in diabetes care. Additionally, we have calculated the ratio between the WTE GP and the WTE nurse spent with diabetics based on the number and type of consultations delivered by GPs and nurses. Table 9.9 below presents this information for each surgery.

We have used the ratio of WTE GP to WTE nurse to group the surgeries into three categories: GP-led (if the ratio is larger than one), nurse-led (if the ratio is smaller than one) and Balanced (if the ratio is approximately equal to one). We have re-mapped the surgeries in terms of their performance results making their system of operation explicit. Analyses of the results taking this factor into account suggest that there is no visible effect on equity, efficiency and effectiveness derived from the three alternative systems of care. There is no indication that the balance between GP and nursing time is a determinant of performance. These results contrast with the results of the study undertaken by Ovhed et al. (2000), where it was found that surgeries where nurses had a more active and independent role in diabetes care performed better in terms of achieving clinical outcomes.

Table 9.9: GPs and nurses role in diabetes care delivery

Surgery	Formal Lead person for diabetes	Professional who runs the diabetes clinics	Ratio of WTE GP to WTE Nurse	At least one GP has formal training in diabetes	At least one practice nurse has formal training in diabetes
1	GP	GP and nurse	1.57	Yes	Yes
2	GP	Nurse	0.85	Yes	Yes
3	GP	Nurse	0.81	No	Yes
4	Nurse	Nurse	0.97	No	Yes
5	Nurse	Nurse	0.82	Yes	Yes
6	GP	GP and nurse	0.85	Yes	Yes
7	GP	No diabetes clinics	1	Yes	No
8	GP	GP and nurse	0.79	No	No
9	No one	No diabetes clinics	1.24	No	Yes
11	Nurse	Nurse	2.53	No	Yes
12	Nurse	Nurse	1.42	No	Yes
13	Nurse	No diabetes clinics	0.78	No	Yes
23	GP	GP and nurse	2.13	Yes	Yes
24	GP	GP and nurse	0.85	No	Yes

9.3.2 Ownership of Diabetes Care

From Table 9.9 we conclude that the formal lead person for diabetes is not necessarily a person with training in diabetes, nor is the person who runs the diabetes clinics. Furthermore, the fact that a particular professional was identified as the formal lead for diabetes does not mean that this professional consults with the diabetics more often. For example, Surgery 11 has identified a nurse as the formal diabetes lead, however the ratio of WTE GP to WTE nurse is 2.53, which suggests that this surgery operates with a GP-led system. This may in fact be sub-optimal because the nurse has formal diabetes training whilst the GP does not. In Surgeries 3, 8 and 24 we have the reverse situation. A GP was identified as the formal lead person for diabetes, however this professional does not have formal training in diabetes. In these three surgeries, the majority of diabetes consultations are delivered by the nurses (WTE nurse is superior to WTE GP).

However, the nurses may not feel that they 'own' the diabetes service delivery because they are not the formal lead for diabetes.

We can identify five surgeries where the formal and the operational roles of diabetes care have been attributed to the same professional: surgeries 1, 4, 5, 13 and 23. In surgery 4 we have one diabetes specialist nurse who is the formal lead person for diabetes, she has an interest and training in diabetes and runs all the diabetes clinics. She works on a P/T basis and only sees diabetics. In Surgery 5, we have two practice nurses who have been identified as the formal lead for diabetes and run the diabetes clinics. In Surgery 13, we have a practice nurse with a special interest and training in diabetes; she delivers most of the consultations to diabetics, despite not running diabetes clinics, and is the formal lead for diabetes. In Surgeries 1 and 23, which are larger surgeries, the diabetes clinics are run both by a GP and a nurse, both professionals have an interest and formal training in diabetes; the GP delivers most of the diabetic consultations and the GP was also identified as the formal lead person for diabetes. All these surgeries have scored higher than average in terms of equity, and apart from surgeries 1 and 13 (which cover highly deprived populations) all of them have scored higher than average in terms of clinical effectiveness.

We certainly would not want to make any generalisations from these results because they are based on a small sample. However, all we want to suggest is that part of the reason why some surgeries don't perform as well as they could may be because they lack a professional (or group of professionals) who feel responsible for a good delivery of diabetes care, takes pride in it and can be identified by the patients as 'the diabetes person'. The existence of such a professional should lead to professional ownership and to the development of a long-term relationship between the patients and the professional, which may contribute to better results both in terms of regular monitoring and in terms of outcomes achievement. The importance of this aspect was also highlighted during the workshops with some of the surgeries.

9.3.3 Other Surgery Related Factors

We will now discuss the effects of other surgery related factors on performance. Table 9.10 presents the data regarding the type of contract that each surgery works with and regarding the training status of each surgery.

Personal Medical Services (PMS) contracts were first introduced in 1998. They provide a more flexible alternative to the General Medical Services (GMS) contracts. PMS contracts were created to fund the delivery of primary care with more innovative approaches. Using these contracts, surgeries can negotiate local arrangements for service provision, including the recruitment of salaried professionals and the commission of services from other organisations. As a result of this increased flexibility, we could expect the performance of surgeries working under PMS contracts to be better than the performance of surgeries working under GMS contracts. The evaluation of the first wave of PMS pilots concluded that some selected PMS sites showed improved access to patients, enhanced intra and inter-professional collaboration and clinical and cost-effective prescribing (The PMS National Evaluation Team 2002: 22). We have grouped the surgeries into two sub-sets according to the type of contracts that they operate under in order to analyse the effects of this factor on performance. No clear effect of this factor on the performance results can be observed.

Table 9.10: Other surgery related factors that may impact on performance

Surgery	GMS versus PMS contracts	Training versus non-training surgery
1	GMS	Non-training
2	PMS	Training
3	PMS	Training
4	GMS	Non-training
5	PMS	Training
6	GMS	Training
7	GMS	Training
8	GMS	Non-training
9	GMS	Non-training
11	PMS	Non-training
12	PMS	Non-training
13	GMS	Non-training
23	PMS	Training
24	PMS	Training

Finally, we have investigated the effect of the training status of the surgeries on their performance. A training practice is expected to keep better records and ensure that a better-organised system of service delivery is in place. For example, Baker (1992) and Van Den Hombergh et al. (1998) concluded that training practices scored higher than average in terms of the use of innovative equipment, delegation of medical tasks and organisation of care. It is therefore interesting to evaluate whether training practices appear to score better in terms of equity, efficiency and effectiveness.

Once again, we separated the surgeries into two sub-sets and looked for patterns in the results. Training status, per se, does not appear to explain the performance results. For example, surgery 4 is a non-training practice and achieved the highest score in terms of equity. This surgery also scores 100% in terms of clinical effectiveness. Another example of a non-training practice that is a relatively good performer is surgery

9. On the contrary, surgery 7 is a training practice, but it appears to have potential for improvement in terms of equity, cost efficiency and clinical effectiveness.

9.4 A Summary of the Impact of the Environmental Variables on the Performance Results

Our analyses regarding the impact of area related factors suggest that surgeries facing a high level of population deprivation may have increased difficulties in achieving a good level of clinical and patient-focused outcomes. In our sample of 14 surgeries, all the surgeries covering areas of high deprivation scored below average in terms of both clinical and patient-focused effectiveness. Furthermore, these were the only surgeries that scored below average on both of these performance criteria.

In this respect, we feel that it would be appropriate to separate the surgeries into three different sets (low deprivation, medium deprivation and high deprivation) and undertake separate comparisons within each set. Alternatively, the level of deprivation could be used as a hierarchical categorical variable, following the methodology proposed by Cooper, Seiford and Tone (2000). According to this methodology, the surgeries facing a high level of deprivation would be evaluated only with reference to other surgeries facing a similar environment. Surgeries facing a medium level of deprivation would be evaluated with reference to surgeries facing a medium and high level of deprivation and surgeries facing a low level of deprivation would be evaluated with reference to all the surgeries.

With regards to patient related factors, we performed two types of analyses. Firstly, we analysed the effect of the level of patient compliance, confidence and understanding on the performance results. We concluded that, for our sample of surgeries, none of these factors appears to have a significant impact on the performance results. Secondly, we analysed the impact of the proportion of elderly patients, newly diagnosed patients, type-1 patients and patients with diabetes related complications on the performance results. It was not possible to identify a significant correlation between any of these variables and the performance results.

With regards to the structures and processes of care, firstly, we investigated the impact of the different models of diabetes care delivery on the performance results. In this respect, we firstly distinguished between surgeries with GP-led diabetes care, surgeries with nurse-led diabetes care and surgeries with a balanced model of care. We concluded that this aspect, per se, does not explain the variability in the performance results. Secondly, we investigated whether the professional with training in diabetes was the formal diabetes lead and delivered the majority of the diabetes consultations. Professional ownership in diabetes care was proposed as one of the factors that may lead to improved performance in terms of equity of access to services and clinical effectiveness. We concluded that in some surgeries there is potential to improve in this aspect. For example, surgery 11 has a nurse with formal training in diabetes, but the majority of the diabetes consultations are delivered by the GP (who does not have training in diabetes). This surgery presents a very expensive type of diabetes delivery, which does not appear to compensate in terms of clinical outcomes, and this may suggest that a greater involvement of the practice nurse in diabetes care delivery could lead to an improvement in cost efficiency and cost effectiveness.

Finally, we investigated the impact of two other surgery related factors on the performance results. Neither the training status of the surgeries or the type of contract with which they work (PMS versus GMS) appears to have a significant impact on the performance results.

Before we continue it is perhaps worth summarising the results for each one of our 14 surgeries, offering some guidance regarding which aspects require particular attention in each surgery. Table 9.11 provides this summary.

9.5 Validation and Implementation of the Results in Practice

We presented the results to each surgery using the trade-off graphs presented earlier. A meeting was requested with each one of the participating surgeries to provide them with the results. The average duration of these meetings was ninety minutes. At the time of submission, meetings had taken place with 8 of the 14 surgeries. In surgeries 7, 8 and 13, all the GPs and nurses took part in the meeting. In surgeries 2, 4, 6, 23 and 24 only the professionals involved in diabetes care took part in the meeting.

In addition to these meetings with the individual surgeries, three further meetings took place. Firstly, a meeting with South Peterborough PCT, including the director of performance and three members of the Board (a GP, a nurse and the manager for primary care services delivery). Secondly, a meeting with the diabetes specialist nurse employed by South Peterborough PCT. She provides extra diabetes care and advice to any surgery that requests her services and is known as 'the trouble-shooting diabetes nurse'. Finally, a meeting with the hospital diabetologist providing services to all the surgeries covered by South Peterborough PCT.

A report with the results and analyses was provided to South Peterborough PCT and to each of the participating surgeries. The identification of each surgery was not disclosed to the PCT or any of the professionals involved in the meetings. It was only during the individual meetings with each surgery that their *own* identification number was disclosed.

Table 9.11: Summary of the performance results for each surgery

Surgeries	Performance criteria that require particular attention	Factors that may partly explain the results
1	Cost efficiency Clinical effectiveness Patient-focused effectiveness	High level of population deprivation. A relatively large proportion of elderly patients.
2	Clinical effectiveness	
3	Cost efficiency Clinical effectiveness Patient-focused effectiveness	High level of population deprivation.
4	Cost efficiency	A system of frequent reviews and intensive treatment operates at this surgery, which leads to higher costs.
5	Cost efficiency	
6	Patient-focused effectiveness	Patient satisfaction requires particular attention. Patients report relatively long waiting times to see a doctor and not enough time to speak with the doctors.
7	Equity Cost efficiency Clinical effectiveness Cost effectiveness	Inadequate system of diabetes management. The surgery does not run diabetes clinics and the patients are not invited for regular reviews. Instead, a system of patient initiated consultations is used.
8	Equity Clinical effectiveness Patient focused effectiveness	High level of population deprivation. Lack of professional training in diabetes care. In this surgery there is no nurse or GP with training in diabetes.
9	Equity	The surgery had a GP vacancy during the year of analysis, which may explain the relatively poor access to services.
11	Equity Cost efficiency	Single GP who delivers most of the diabetic consultations.
12	Equity Cost efficiency	Single GP who delivers most of the diabetic consultations.
13	Clinical effectiveness Patient-focused effectiveness	High level of population deprivation.
23	Patient-focused effectiveness	Patient education requires particular attention. The level of patient understanding is lower than average.
24	Equity Cost efficiency	The F/T diabetes specialist nurse working at this surgery follows a policy of intensive treatment with medication, which may explain the relatively high prescription costs incurred.

9.5.1 Assessing the Reputational View of Performance

In order to assess the 'reputational view' of performance, we asked those who took part in the PCT meeting, the trouble-shooting diabetes nurse and the hospital diabetologist to place the 12 surgeries from South Peterborough PCT in three blank performance trade-off graphs. Based on their responses, we concluded that surgeries 2, 4, 6 and 9 were identified as good performers by all professionals. The perception regarding the other surgeries tended to vary between professionals. The interesting aspect to notice is that a perception of the potential trade-offs between these criteria was not always evident. Surgeries 4, 6 and 9 are in fact the three surgeries that appear as robust performers. Surgery 2 is an interesting case because despite the fact that it scores very highly in terms of equity and efficiency, it scores below average in terms of clinical effectiveness, which was a cause of surprise to most professionals involved in the workshops.

We also asked each surgery to make a self-assessment of their performance in terms of the different performance criteria. In most cases, the surgeries perceived their relative performance to be better than what was shown by the results, exception made to surgeries 4, 13 and 23, which appeared to have a good idea regarding their relative performance in the different criteria. An aspect of interest to notice is the fact that the surgeries had never considered labour as a cost of diabetes care and were therefore very surprised to see it included in the analyses. In their internal audits, the only cost that was considered was the cost of prescriptions.

9.5.2 Using the Results in Practice

Both the PCT and the individual surgeries felt that the results of the research project were very useful in terms of pointing towards areas that needed to be improved. The results from the patient questionnaire were particularly useful in identifying aspects of care that needed to be improved in each surgery, as well as recognising some of the positive aspects. The wide variation in the results, the implicit trade-offs between some of the performance measures, and the impact of area deprivation were recognised as important issues that needed to be further investigated.

We believe that the fact that the performance assessment framework and the patient questionnaire were discussed with all the surgeries prior to the data extraction was an important part in the process. Furthermore, keeping anonymity of the results, together with an emphasis on formative evaluation provided the elements for an open discussion of the results. We believe that this is essential in order to ensure that the performance results are taken seriously and acted upon by each provider. Below, we provide some specific examples of the actions taken by each surgery after the discussion of the results.

The formal lead for diabetes in Surgery 2 (a GP) was surprised to discover that his surgery scored relatively poorly in terms of clinical effectiveness. The good results in terms of equity and cost efficiency were pointed out as positive aspects, and it was suggested that this might partly explain the low level of clinical effectiveness achieved. The GP said that his surgery follows a policy of high coverage, aiming to review most of their diabetics yearly, and that this is their first objective in terms of diabetes care. Nevertheless, the need to investigate the relatively poor level of outcomes achievement was recognised. In terms of the results from the patient questionnaire, particular relevance was given to the fact that 50% of the patients who answered the questionnaire said that they were not always satisfied with the diet advice received at the surgery.

The diabetes specialist nurse in Surgery 4 was particularly pleased with the results, as she felt that her work and dedication were shown in the study. The nurse acknowledged that the location of the surgery in a non-deprived area was an important factor contributing to the relatively good performance results. Furthermore, her established autonomy and professional pride in diabetes care delivery were suggested as essential elements in the achievement of good results. From the results of the patient questionnaire, the nurse decided to improve in terms of the quality of the educational materials provided to the patients. The usefulness of the educational materials was the quality item where this surgery scored lowest. In this respect, one of the walls in the patients waiting area was covered with colourful and informative posters regarding the importance of diabetes control, including the results from some research studies. Furthermore, new leaflets were produced and are now being distributed to the diabetic patients.

Surgery 7 was surprised to find that its performance results were lower than average in most performance dimensions. This surgery does not operate diabetes clinics. The justification for this is that these clinics tend to limit the flexibility of the appointment times and appear to be less convenient for patients. However, the results from the study prompted them to discuss different models of diabetes care delivery in order to improve in terms of equity and effectiveness. Interest in meeting the specialist diabetes nurse from surgery 4 (one of the peers for surgery 7) was expressed, and contacts were made to arrange a joint workshop. The specific comments from the patient questionnaire were used to improve the services. Four essential issues were identified: (1) the surgery is very busy and patients feel that they never see the same doctor, which compromises continuity of care; (2) poor communication between health care professionals and conflicting advice given to patients; (3) the patients are not invited for an annual review. Instead, the patients have to take the initiative and make their own appointments; (4) a number of patients felt that they do not have enough time to talk with the doctors.

The diabetes specialist GP in Surgery 23 was very pleased with the results. However, his suspicions regarding a relatively low level of patient satisfaction were confirmed and therefore he decided to target this area as an area for improvement in the near future. Five essential aspects that can be improved were identified: (1) patronising attitude of some nurses and some of the administrative staff; (2) poor written education materials, particularly in terms of diet and effects on diabetes control; (3) poor initial consultations to newly diagnosed patients; (4) some patients felt that their concerns are not being listened to; (5) difficulties in making appointments with the diabetes specialist GP between reviews.

Surgery 24 recognised that despite the fact that they achieve a high level of clinical effectiveness; their system of diabetes care delivery still needs to improve in terms of equity and in terms of cost efficiency. All the diabetics that had not received an annual review were identified and a letter was sent inviting them for a full annual review. Furthermore, given that Surgery 23 is one of the peers to Surgery 24, the diabetes specialist nurse was very interested in meeting with the diabetes care team from Surgery 23 in order to exchange ideas regarding diabetes care delivery. Contacts were established to arrange a joint workshop.

At a more strategic level, the PCT is particularly interested in improving the levels of equity and cost efficiency. In order to increase the proportion of patients receiving a full annual review, the services of a health care assistant were offered to the surgeries that scored below average in terms of equity and cost efficiency.

This research project is still ongoing and other meetings with South Peterborough PCT will be scheduled to discuss the results in more depth and to assess other ways in which these results can be used in practice.

9.6 What Role for DEA in Formative Evaluation in Primary Care?

The purpose of this study was to use DEA in the formative mode in order to better understand the reasons behind differences in performance in primary care and in order to identify strategies for performance improvement in practice. The small number of surgeries that has been compared placed strong limitations regarding the type of analyses that could be made and regarding the robustness of the conclusions taken. Nevertheless, it is worth making a tentative assessment regarding the usefulness of the results in practice and the limitations faced.

The most important aspect to notice relates to the great deal of variability found in the performance results amongst a small sample of surgeries. Even after narrowing the analyses down to a speciality, using a performance assessment framework developed with the providers, and considering the efforts made to extract data of high quality, a great deal of variability was still found in most of the performance criteria. Part of the variability can be explained by the small numbers of patients involved and by the fact that a single measurement was used for each variable. Furthermore, the level of population deprivation was found to explain part of the variability, especially in terms of effectiveness. However, there remains a significant amount of variability that could not be adequately explained and that needs to be carefully interpreted.

Our aim has been solely exploratory and a great deal of care has been taken to prevent the misuse of the results. In this respect, no rankings or targets have been produced and anonymity of the surgeries has been kept. We have to recognise that the DEA results have not answered many questions. Instead, they raised a number of questions, whose answers only those involved in the delivery of the services can appropriately provide. This is the reason why we believe that the establishment of learning networks between the different surgeries can bring interesting insights.

Nevertheless, despite our formative aim, the potential unintended consequences that may result from any performance assessment exercise cannot be ignored (Smith 1995). If the potential benefits resulting from a performance assessment exercise do not outweigh the costs, then one must ask whether the exercise is worth the efforts (Smith 1990, Power 1997). Answer to this question requires careful investigation of the effects of the study in practice, which is not easily accomplished. This difficulty is exacerbated because the use of formal mechanisms of performance assessment can result in the loss of informal mechanisms of quality assurance, such as personal pride, professional commitment (Smith 1992) and trust (Davies and Mannion 2000). In this respect, a combination of formal and informal mechanisms for performance assessment may be more advantageous (Goddard, Mannion and Smith 1999, Mannion and Smith 2000). The use of DEA, a formal mechanism, in a formative and exploratory mode, can harness the benefit of both formal and informal approaches.

9.7 Conclusion

In this chapter we have analysed the impact of the environmental variables on the performance results. We have also investigated the impact of different structures and mechanisms on the performance results. The usefulness of the results in practice was discussed and strategies for performance improvement were outlined for some of the participating surgeries.

We analysed the impact of area deprivation and surgery location, and concluded that surgeries covering populations with a high level of deprivation showed lower than average results in terms of effectiveness. Furthermore, we analysed the impact of patient cooperation and patient characteristics on the performance results. None of the variables considered was found to correlate significantly with the performance results.

Finally, our analyses of the performance results together with the interviews undertaken with the participating surgeries suggest that one way to improve the quality of diabetes service delivery might be to attribute it to one professional (or group of professionals) with an interest and training in diabetes care. This professional is given the responsibility to ensure that all registered patients get all the essential services once a year. If nobody else has this responsibility, clear accountability can be established and a sense of professional ownership is also achieved. If this model is to be effective in terms of outcomes achievement, the professional needs to have some autonomy regarding diabetes prescribing, and good communication with all the other professionals consulting diabetics has to be ensured. Above all, a long-term relationship of trust needs to be established between the professional and the patients.

10 Conclusions

10.1 Summary and Contributions of this Study

The purpose of this thesis was to address issues raised by performance assessment in primary health care, in order to develop a comprehensive framework to compare primary care providers. In particular, considering performance as multidimensional, this study has addressed the question of conflicting objectives in health care provision, and attempted to investigate some of the potential trade-offs between different performance criteria. A discussion about ways to achieve an adequate balance between several performance criteria was also included.

There are two aspects in this study, which distinguish it from others in this research area: (1) the framework used and (2) the use of this framework. This study can be characterised as a pluralistic evaluation, given that it has focused on the comparison of primary care providers using multiple criteria and given that it used several methods to collect and analyse the data. Furthermore, given its objectives and the methodology used, it can be characterised as a formative evaluation study.

A review of the studies that have been published comparing primary care providers was undertaken. Based on this review we identified some issues that needed to be addressed in order to develop a comprehensive performance assessment framework. Firstly, the use of PIs was discussed and it was argued that they tend to be inadequate, as they do not establish a link between the resources used, services delivered and outcomes achieved. The use of parametric techniques was recognised as a suitable alternative to PIs, but the difficulties in specifying an appropriate functional form for the transformation processes in primary care were an obvious obstacle. In this respect, DEA was proposed as a suitable technique to compare primary care providers because it can manage the existence of: multiple inputs, multiple outputs and multiple outcomes; and does not require the specification of a functional form.

Previous studies that have used DEA to compare primary care providers were reviewed and it was argued that they have focused mostly on structure and outputs, without taking into account the achievement of outcomes, which is the ultimate aim of primary care delivery. The potential for unintended consequences resulting from the publication of performance results, particularly in the form of league tables, was rehearsed and it was suggested that DEA can be most useful if it is used in a formative and exploratory way. In this respect, the importance of involving the decision-makers during the full process of evaluation was highlighted, both to provide the researcher with a better understanding of the context and to increase the confidence of the decision-makers in the results.

Based on the issues identified in the literature review, a conceptual framework was developed in order to compare GP surgeries (Figure 4.3 presented in page 84). This framework was developed in collaboration with a sample of PCG/Ts and its aims were to establish a link between local needs, resources used, services delivered and outcomes achieved in primary care delivery. By focusing on the relationships between these elements, different performance assessment criteria were proposed: equity, technical efficiency, cost efficiency, service effectiveness and cost effectiveness. Three methods were proposed to measure equity: (1) resources over needs; (2) services delivered over needs; (3) outcomes over needs. The use of DEA was suggested to measure technical efficiency, cost efficiency, service effectiveness and cost effectiveness. The approach emphasised the use of several inter-connected DEA models to measure performance as opposed to the use of a single all-embracing DEA model.

The feasibility of applying the conceptual framework developed to compare GP surgeries in terms of their delivery of primary care as a whole was then discussed. The absence of appropriate data for outcomes of primary care and the lack of well determined models of primary care delivery that work within the production metaphor, meant that the comparison of GP surgeries in terms of their delivery of primary care as a whole would not be feasible at this stage. The existence of clear protocols regarding the delivery of diabetes care and the availability of appropriate data suggested that a formative evaluation in this context would be feasible. Furthermore, given the rising costs of this condition and the evidence regarding varying standards of practice across different providers (The Audit Commission 2000b), it was felt that a formative evaluation of primary diabetes care delivery would be very timely. Previous studies focusing on primary diabetes care delivery were reviewed and it was argued that most studies failed to establish a link between local needs, resources used, services delivered and outcomes achieved.

The conceptual framework for primary care evaluation was therefore applied to diabetes care at the level of GP surgeries. Workshops were undertaken with the participating PCG/Ts in order to identify the specific measures for local needs, inputs, outputs and outcomes in primary diabetes care delivery. Based on this framework, a comparison of a sample of GP surgeries was then undertaken. The objective of the comparison was a formative one, which focused on (1) the investigation of the relationships between the different performance criteria and (2) the identification of criteria in which the performance of each surgery could be improved. In this respect, this study has provided some interesting insights regarding the use of DEA as a tool for performance improvement in practice.

Firstly, in agreement with available literature (see for example, Smith 1997; Garcia et al. 1999 and Parkin and Hollingsworth 1999), we found that DEA results can be greatly influenced by the model used. For example, adding a new input or output may lead to different results, conclusions and actions. In this respect, we would like to re-emphasise the importance of *remembering* that the performance results are a function of the model we have used, and that this model is only one out of several plausible specifications. It is in this respect that the use of theories and evidence regarding the links between inputs, outputs and outcomes becomes crucial. Additionally, the involvement of the providers in the development of the models provides another source of model validation.

Secondly, we found that using input prices leads to more meaningful estimates of efficiency, given that surgeries are not allowed to choose freely an input weight structure that may lead to non-realistic technological trade-offs. Furthermore, the decomposition of efficiency into technical and allocative efficiency, together with an analysis of the optimal input mixes helps in the investigation of the sources of inefficiency. Whilst some surgeries suffer from technical inefficiency and should be able to radially reduce all the inputs, others appear to use a non-optimal mix of resources.

Thirdly, we found that DEA can also be very useful in measuring both clinical effectiveness and cost effectiveness. An analysis of the effectiveness results provides another perspective into performance assessment, which is essential if we are to use the results for performance improvement.

Fourthly, we found that complementing the DEA results with an analysis of equity of access to services is essential if we are interested in assuring a minimum standard of quality of access to services. Some of the surgeries that were deemed efficient fail to provide the essential services to a significant proportion of their patients.

Finally, we found that reporting the results to the providers by using trade-off graphs was extremely helpful. This approach does not emphasise the ranking of the providers, nor does it emphasise the performance scores, or performance targets. Instead, it shows that the performance results change with the criterion used. Furthermore, it shows that in most cases, learning can happen in two directions. For example, Surgery A can learn from surgery B in terms of efficiency, whilst surgery B can learn from surgery A in terms of effectiveness. We feel that this is a very useful way to initiate a discussion of performance improvement strategies, as most providers were happy to discuss their way of working as long as they had not been characterised as 'bottom of the league performers'.

The use of trade-off graphs also provided interesting insights regarding the relationships between the different performance criteria. We investigated the relationships between equity and cost efficiency, cost efficiency and clinical effectiveness, equity and clinical effectiveness and clinical effectiveness and patient-focused effectiveness.

Firstly, for the sample of surgeries studied, we could not find a significant trade-off between equity and cost efficiency. On the contrary, a strong positive correlation between these two measures was identified for a group of surgeries. This suggests that a high level of service accessibility does not have to compromise cost efficiency. It appears to be possible to develop systems of diabetes care delivery that are cost efficient and simultaneously respect standards of accessibility. When we consider the fact that our analyses are based on a small sample, we feel that it would be interesting to further investigate this relationship using a larger sample of surgeries. Furthermore, we feel that it would be interesting to extend this type of analyses by measuring equity of access to services focusing on factors such as the socio-economic group and ethnicity of patients.

Secondly, in terms of the relationship between efficiency and effectiveness, our results did not show a significant relationship between these two performance measures, despite the fact that during the workshops, most health care professionals hypothesised the possibility of this trade-off. Furthermore, our results did not confirm the theory proposed by Schinnar (1993) regarding greater effectiveness levels for moderate levels of efficiency. We identified a group of surgeries, which perform very highly in terms of both efficiency and effectiveness. The investigation of this relationship using larger samples is therefore required in order to reach firm conclusions.

Thirdly, we investigated the relationship between equity and effectiveness. For our sample of surgeries, it was not possible to identify a significant trade-off between these two measures. Nevertheless, we feel that this trade-off is worth further investigation, given the potential dysfunctional incentives that may result from the implementation of the new GP contract. The establishment of performance related pay based on outcome targets, as proposed in the new GP contract (British Medical Association 2003), might lead to a decrease in the level of fairness of access to services. There is a danger that professionals may start to focus on the patients that appear to cooperate and those that appear to lead more effectively to the achievement of the performance targets, *at the expense of* more difficult cases, for whom only a box ticking exercise will be carried out. Additionally, exercises of patient re-coding can take place in order to report better outcome results. Furthermore, there is a danger that some selection bias could take place at new patient registration. Hence the importance of investigating equity when assessing primary care providers.

Finally, we investigated the trade-off between clinical effectiveness and patient-focused effectiveness. Whilst clinical effectiveness focuses on the achievement of clinical outcomes, patients focused effectiveness is concerned with the achievement of patient related outcomes such as patient satisfaction, patient understanding and patient confidence. No clear relationship between these two performance criteria was identified. However, it was interesting to notice that the surgeries that performed above average in terms of both clinical and patient focused effectiveness, all performed below average in terms of cost efficiency. Furthermore, a significant negative correlation was identified between cost efficiency and patient focused effectiveness, which may suggest that higher levels of patient understanding, confidence and satisfaction may be associated with higher levels of spending. This takes us into a crucial issue in performance assessment in primary care: the role of the patients as co-producers of care. Patient characteristics and willingness to cooperate can be expected to have an impact on the performance of the surgeries.

We investigated the impact of both area related factors and patient related factors on the performance of the surgeries. In terms of area related factors, we investigated the impact of different levels of deprivation and the impact of different locations. With regards to patient related factors we investigated the impact of patient non-compliance, understanding and confidence; and also investigated the impact of the proportion of patients with diabetes related complications, elderly patients, newly diagnosed patients and type-1 patients.

In consistency with the results of other studies (Roper et al. 2001; Acheson Report 1998; Urwin et al. 1996; Connoly and Kesson 1996; Kelly et al. 1993), our study suggests that surgeries covering highly deprived populations may find it more difficult to achieve good clinical and patient-focused outcomes in diabetes care delivery. In this respect, it may be appropriate to develop different performance assessment frameworks for surgeries covering highly deprived areas. These surgeries should be subjected to different outcome targets, and should only be compared with each other. Furthermore, it may be necessary to develop different models of diabetes care delivery in order to increase the level of effectiveness. Strategies characterised by a wider action, aiming to reduce the discrepancies in terms of area deprivation would also be appropriate.

In terms of our analysis of the impact of patient related factors on the performance results, we concluded that none of the factors individually explained the variability obtained in the performance results. The performance variability is likely to be the result of a complex combination of multiple factors.

10.2 Implications for Health Care Policy

We have attempted to draw some implications for primary health care policy and diabetes care delivery from the analysis of the performance results and from the workshops undertaken with the healthcare professionals.

Considering the difficulties we faced in extracting the necessary data for the performance analysis, we would suggest that a great deal of improvement in the data recording and data retrieval practices still needs to take place before an extensive formative evaluation exercise in primary care can be implemented. This need for improvement is evident in terms of several aspects: (1) the electronic entry of information regarding consultations and procedures taking place at the surgery, (2) better coordination of information between the surgeries and other providers by electronic recording, (3) the standardisation of the codes used by the professionals in different surgeries and (4) better familiarisation of the professionals with simple ways to extract relevant data for audit and quality control.

Further issues regarding inadequate data availability relate to the difficulty in extracting data for the costs of laboratory tests. Currently, each surgery is not charged in proportion to the number of laboratory tests requested. A block contract exists between the PCTs and the laboratories, and payment is independent of the number of tests requested. With a system driven by strong outcome targets, and without a monitoring mechanism, we think that there is a clear danger of over-testing and resource wasting. Furthermore, our study suggests that due to the poor coordination between services, there is a danger of service duplication across different providers. For example, we found that a significant proportion of patients had their eyes checked and their blood tested at several different providers. Whether or not these services were necessary, we cannot say, but a better coordination between the providers could prevent such duplication from happening. Improved coordination between providers, with electronic recording of the services provided by the hospital and other local providers can also ensure that all patients receive the services that they need.

A further issue that was identified during our study relates to the difficulty of extracting data regarding the costs involved in primary diabetes care. Speciality based costing is not undertaken in primary care, despite the fact that speciality based frameworks for service delivery and for performance assessment are currently being implemented in the UK. We feel that if speciality based targets are going to be imposed on the surgeries, it is essential to take into account the costs involved in their achievement. If appropriate additional funding is not provided, the surgeries may attempt to meet diabetes related targets by 'shifting' resources from other areas of primary care delivery where targets have not been imposed.

Furthermore, we feel that the strategy of speciality based service delivery in primary care can have other dysfunctional consequences. The existence of several non-integrated National Service Frameworks without a global framework that integrates all the specialities and draws from the synergies between them can lead to a fragmented delivery of primary care. We feel that this may lead to a disruption in the personal and continuous relationship between primary health care professionals and patients. This continuous relationship is essential to build trust and is one of the founding bases of primary care. We feel that the consequences of disrupting this relationship should be further investigated before moving into a more specialised primary care delivery system. Furthermore, a system characterised by a specialist delivery of primary care services may lead to the potential loss of skills of GPs in terms of the diagnosis and management of certain diseases.

Patient responsiveness and patient choice are two of the stated objectives for the NHS (The Department of Health 1997). Our study suggests that the level of patient responsiveness in diabetes care delivery could be improved. A significant number of the patients who answered the patient questionnaire expressed strong dissatisfaction regarding the large number of appointments they have to make and the inconvenience of the appointment times. The multiplicity of appointments causes great costs to patients, who find that they have to take several days off work in order to attend the appointments. Furthermore, it creates a stigma around this condition, when in fact a large proportion of diabetics lead a healthy life. For a chronic condition that affects a large number of working patients, it is urgent to develop new models of service delivery more adapted to patients' needs and lives. For example, a 'one-stop' approach for reviews and tests combined with follow-up telephone consultations could both save money and increase patient responsiveness.

Another strategy outlined by this Government regarding the management of chronic conditions is to increase patient empowerment by giving patients a better understanding of their condition (The Department of Health 2001b). This strategy assumes that patients both want and have the ability to be empowered and to take control of their condition. The results from our patient survey reveal that 99.2% of the patients think it is important to have a good understanding of diabetes and its management. However, we found that only 49% of the patients were satisfied with the education materials provided by the surgeries. There is a need for more innovative approaches. Most of the surgeries studied only offer leaflets to the patients. However, leaflets are very difficult to update and are difficult to tailor to the different levels of patient need and understanding. It is common for patients to be given leaflets covering both types of diabetes without an appropriate explanation regarding which type of diabetes they have and which information is therefore relevant to them. A computer could be installed in the waiting area of the surgery and patients could have a chance to fill in a questionnaire about the quality of the services and about their level of understanding, as well as to consult their test results. The computer could then offer patients some up to date, personalised advice, regarding their diabetes management. All of this could be done while patients wait to see the nurse or the GP. Furthermore, a website could be created to provide up to date information to the patients who have access to the Internet.

10.3 Limitations of this Study and Suggestions for Further Research

This study suffers from several limitations. Below we discuss these limitations and suggest ways in which this research can be extended.

Firstly, it is very difficult to draw robust conclusions from a sample of 23 surgeries. The investigation of a larger sample of surgeries would be required in order to obtain more reliable results and to establish more reliable conclusions. The use of a larger sample would also have allowed separate analyses by surgeries covering areas with different levels of deprivation and the inclusion of indicators regarding patients' characteristics.

Secondly, this study represents a snapshot of performance. A dynamic analysis of data for several years could provide very interesting insights into the evaluation of performance, as well as provide us with greater confidence regarding the robustness of the results. If data for several years was available an evaluation including indicators of change in the level of outputs and outcomes could also be undertaken.

Thirdly, we did not take into account the influence of stochastic factors on performance. It can certainly be expected that part of the variability encountered is the result of natural stochastic factors. If a larger sample of providers was available, SFA could be used to separate between stochastic variability, the effects of environmental factors and pure inefficiency and ineffectiveness (Fried et al. 2002).

Fourthly, it would have been useful to undertake a longer qualitative investigation into the usefulness of the performance results in practice. A posterior investigation regarding the effects of this study in practice, with particular relevance to the usefulness of DEA information about peers and targets, could provide interesting insights. Unfortunately, due to limitations of time both from the part of the researcher and the part of the health care professionals it was not possible to undertake any further workshops before submission. This is the subject of future research.

Fifthly, given the small number of diabetics registered with most of our surgeries, it was not possible to investigate the equity of utilisation of services across different groups of patients within each surgery. An assessment of the equity of utilisation across different socio-economic groups and across different ethnic groups should be included in a future study undertaken in this research area.

Finally, the difficulties in estimating the inputs that relate solely with diabetes care delivery meant that some assumptions had to be made regarding the level of resources usage. All the consultations with diabetics were included as an input of diabetes care and an average duration of consultations across surgeries had to be assumed because some surgeries were unable to provide us with data for the duration of their consultations.

The difficulties encountered in terms of obtaining appropriate data for performance assessment in primary care delivery suggest that the successful application of DEA in this research area is still a challenge. Nevertheless, it is our conviction that DEA is a strong analytical technique upon which to build. However, one of the premises of this thesis is that DEA has to be broadened to address problematic situations such as those faced by today's NHS in order to contribute to performance improvement in practice. One part of this broadening entails making values and assumptions explicit before engaging in the performance measurement exercise. The other part of the broadening relates to the need to move away from the 'black box' type of evaluation, by incorporating the DEA exercise into case studies and context-driven research projects in order to facilitate the implementation of the results in practice. As pointed out by Goddard, Mannion and Smith (1999) the combined use of 'soft' and 'hard' information plays a crucial role in performance assessment in the health care sector given the complexity of the 'production process'.

By using DEA in a formative mode, this research has contributed to a discussion regarding the usefulness of this technique for performance improvement in practice. The existence of multiple, incommensurable and potentially conflicting values in public services delivery was acknowledged and a framework was developed in order to account for potential trade-offs between the different criteria. Future research studies should investigate ways in which DEA can be combined with other methodologies in order to better inform decision-making and policy making in public services delivery.

> "No amount of guidance from the NHS Executive or hectoring by politicians can substitute for a drive to improve performance that comes from within and is acknowledged and valued by those steering the process of change."
>
> (Ham 1999: 1492)

Best Copy Available

P257 — P288

Appendix A High Level Performance Indicators for Health Authority

High Level Performance Indicators (Health Authority Level)
I. Health Improvement
Deaths from all causes (for people aged 15-64)
Deaths from all causes (for people aged 65-74)
Cancer registrations
Deaths from malignant neoplasms
Deaths from all circulatory diseases
Suicide rates
Deaths from accidents
II. Fair Access
Surgery rates
Size of inpatient waiting list per head of population (weighted)
Adults registered with an NHS dentist
Children registered with an NHS dentist
Early detection of cancer
III. Effective Delivery of Appropriate Health Care
Disease prevention and health promotion
Early detection of cancer
Inappropriately used surgery
Surgery rates
Acute care management
Chronic care management
Mental health in Primary Care
Cost effective prescribing
Discharge from hospital
IV. Efficiency
Day case rate
Length of stay in hospital (case-mix adjusted)
Unit cost of maternity (adjusted)
Unit cost of caring for patients in receipt of Specialist mental health services (adjusted)
Generic prescribing
V. Patient / Carer Experience of the NHS
Patients who wait less than 2 hours for emergency admission (through A&E)
Patients with operation cancelled for non-medical reasons
Delayed discharge from hospital for people aged 75 or over
First outpatient appointments for which patient did not attend
Outpatients seen within 13 weeks of GP referral
Percentage of those on waiting list waiting 12 months or more
VI. Health Outcomes of NHS Health Care
Conceptions below age 16
Decayed, missing and filled teeth in five year old children
Adverse events / complications of treatment
Emergency admissions to hospital for people aged 75 and over
Emergency psychiatric re-admission rate
Infant deaths
Survival rates for breast and cervical cancer
Avoidable deaths
In-hospital premature deaths

Source: NHS Executive (1999).

Diabetes Services Questionnaire

<Patient's Name>
<Patient's Address 1>
<Patient's Address 2>
<Patient's Address 3>
<Patient's Post Code>

Dear <Patient's name>,

Your GP surgery is asking people with diabetes what they think about the services provided, and we would be grateful if you could help by filling this questionnaire. **This is a confidential, anonymous survey**.

The answers to this questionnaire will be analysed by an independent researcher from the University of Warwick, studying the quality of diabetes service delivery. **What you tell us will help doctors and nurses to improve diabetes services locally**.

We expect that this questionnaire will not take you more than 10 minutes to complete.

Please feel free to use any extra paper for your comments and suggestions for the improvement of the services.

We would appreciate if you could answer all the questions and return the questionnaire in the sealed FREEPOST envelope provided.

Please write in BLOCK CAPITAL LETTERS.

The results from the survey will be fed-back to you via your Practice.

Thank you very much for your help.

<u>Section A</u> – 1) About the services you have received <u>in the past 12 months</u>.

We would appreciate if you could check your records or confirm with someone in your family the number of times you have received each service both at the GP surgery and somewhere else.

- Please tick a box correspondent to the number of times you received each service in the last 12 months, or fill in the empty square with the number of times, if it is more than 3 times;
- Please separate the number of times you had it at the GP surgery and the number of times you had it outside the GP surgery.

<u>Services received during the past 12 months</u>	**Number of times you had it done <u>at the GP Surgery</u>**	**Number of times you had it done <u>outside the GP Surgery</u>**
a. Eyes check (with drops put in the eyes)	0 1 2 3 ☐	0 1 2 3 ☐
b. Legs and feet check (without shoes/socks)	0 1 2 3 ☐	0 1 2 3 ☐
c. Blood Pressure check	0 1 2 3 ☐	0 1 2 3 ☐
d. Long term Blood Sugar level check (HbA1c)	0 1 2 3 ☐	0 1 2 3 ☐
e. Blood cholesterol check	0 1 2 3 ☐	0 1 2 3 ☐
f. Blood/ urine tests to check if the kidneys are working well	0 1 2 3 ☐	0 1 2 3 ☐
g. Consultation <u>with a dietician</u>	0 1 2 3 ☐	0 1 2 3 ☐
h. Consultation <u>with a doctor</u> to review the diabetes treatment	0 1 2 3 ☐	0 1 2 3 ☐
I. Consultation <u>with a nurse</u> to review the diabetes treatment	0 1 2 3 ☐	0 1 2 3 ☐

2) If you have not received some of these services during the last 12 months, or if you did not find some services useful, please explain why:

____

____

____

Section B - **About your Understanding of Diabetes and its Management**

1. In relation to your diabetes, how much do you think you understand about each one of these aspects? Please tick one box for each statement.

Aspects of diabetes	I understand nothing	I understand very little	I understand enough	N/A
a. Objectives of treatment	☐	☐	☐	☐
b. Diet requirement and meal planning	☐	☐	☐	☐
c. Benefits of regular exercise	☐	☐	☐	☐
d. Improvements in life style - effects of smoking and alcohol intake (If you not smoke or drink, tick N/A)	☐	☐	☐	☐
e. Link between food intake, exercise and tablets or insulin	☐	☐	☐	☐
f. How to cope with diabetic emergencies (e.g. high or low blood sugar)	☐	☐	☐	☐
g. Signs, symptoms, and prevention of complications of feet, eye and heart	☐	☐	☐	☐
h. Effects of an illness like flu on your diabetes	☐	☐	☐	☐
I. Measuring your blood sugar, meaning of values and actions to take	☐	☐	☐	☐
J. Insulin injection (If insulin is not required, tick N/A)	☐	☐	☐	☐

•

Section B – Continuation

2. How important is it for you to have a good understanding of diabetes and its management?

Not important ☐ Important ☐ Very Important ☐

3. Overall, how confident do you feel in taking good control over your diabetes?

Not confident ☐ Confident ☐ Very Confident ☐

4. Where do you find the most helpful information regarding your diabetes? (Please tick all the relevant options)

GP Surgery ☐ Hospital ☐ Library ☐ Internet ☐ Diabetic associations ☐

Other (please specify) ______________________________

5. What type of information do you find most helpful? (Please tick all the relevant options)

Individual consultation with a doctor ☐ Individual consultation with a nurse ☐

Presentation to a group of patients ☐ Discussions with other patients ☐

Books ☐ Leaflets ☐ Magazines ☐ Videos ☐ TV Programmes ☐

Other (please specify) ______________________________

6. Who do you think has helped you most regarding your diabetes management and control?
(Please tick all the relevant options)

GP ☐ Specialist Doctor ☐ Nurse ☐ Family and Friends ☐

Other Patients ☐ Social Services ☐

Other (Please specify)

• ______________________________

<u>Section C</u> - About the Diabetes Care You Received <u>at the GP Surgery</u> in the Past 12 Months

1.1. Please tick one box for each statement, depending on whether you think that it corresponds to the truth 'Always', 'Most times', 'Sometimes' or 'Never'.

1.2. Please <u>say how important each of these aspects of care are for you</u>, by choosing a number for each statement, with: **1 = Very important; 2 = Important; 3 = Not important**.

Statement	Always	Most times	Sometimes	Never	Importance
a. It is easy for me to get health advice at the GP Surgery <u>when I need it</u>	☐	☐	☐	☐	1 2 3
b. The time I have to wait for an appointment with the <u>doctor</u> is acceptable	☐	☐	☐	☐	1 2 3
c. The appointment days and times given at the surgery are convenient to me	☐	☐	☐	☐	1 2 3
d. The <u>doctors and nurses</u> involve me in the decisions about the treatment of my diabetes	☐	☐	☐	☐	1 2 3
e. The <u>doctors and nurses</u> explain the importance of following their advice, in a way that I understand	☐	☐	☐	☐	1 2 3
f. The <u>doctors and nurses</u> are very caring and supportive	☐	☐	☐	☐	1 2 3
g. I feel that the <u>doctors</u> have enough time to talk/ listen to me regarding my diabetes	☐	☐	☐	☐	1 2 3
h. I feel that the <u>nurses</u> have enough time to talk/ listen to me regarding my diabetes	☐	☐	☐	☐	1 2 3
i. I receive <u>consistent advice</u> about my diabetes and treatment	☐	☐	☐	☐	1 2 3
j. I am satisfied with my <u>diabetes treatment</u>, including the medication I have to take and its effects on the quality of my life	☐	☐	☐	☐	1 2 3
k. I am satisfied with the <u>foot care</u> that I receive	☐	☐	☐	☐	1 2 3
l. I am satisfied with the <u>diet advice</u> that I receive	☐	☐	☐	☐	1 2 3
m. The surgery provides me with <u>useful educational materials</u> regarding diabetes	☐	☐	☐	☐	1 2 3

Section C – Continuation

2. Have you visited your GP surgery during the last 12 months for any service other than the diabetes services?

Yes [ ]   No [ ]

3. If the answer to the last question is Yes, how would you rate the quality of the other services provided at your GP surgery?

Very Good [ ]   Good [ ]   Reasonable [ ]   Poor [ ]   Very Poor [ ]

Section D - About you

1. **How old are you?**

Under 20 [ ]   21 – 40 [ ]   41 – 50 [ ]   51 – 65 [ ]   66 – 75 [ ]   Over 75 [ ]

2. **You are:** Male [ ]   Female [ ]   3. **Do you smoke?** Yes [ ]   No [ ]

4. **To which ethnic group do you belong?** ______________________

5. **Which one of these categories best describes your current situation?**

In paid employment [F/T] [P/T]   **What is your job title?** ______________________

Unemployed [ ]   Retired [ ]   Looking after the family/ home [ ]   Full-time student [ ]

Long-term sick or disabled [ ]   Other (please specify) ______________________

6. **What is your level of education** (e.g. Degree, O-level, etc...)? ______________________

7. **How old were you when your diabetes was diagnosed?**

Under 20 [ ]   21 – 40 [ ]   41 – 50 [ ]   51 – 65 [ ]   66 – 75 [ ]   Over 75 [ ]

8. **How do you control for your diabetes?**

Diet only [ ]   Diet and Tablets [ ]   Diet and Insulin [ ]   Diet, Tablets and Insulin [ ]

9. **Please use this space to make comments regarding the diabetes services and suggestions for their improvement (Feel free to write overleaf, if necessary):**

____

____

- **Thank you for your time.**

Please use the freepost envelope and post the questionnaire ASAP.

Appendix C Copy of Ethics Approval Letter

MULTI-CENTRE RESEARCH ETHICS COMMITTEE

RESPONSE FORM

DETAILS OF APPLICANT:

1. **Name and address of Principal Researcher:**

Ms Carla Amado
University Lecturer
Warwick Business School
The University of Warwick
Coventry
CV4 7AL

2. **Title of project:**

Linking Costs and Outcomes of Diabetes care Towards a more effective service Delivery

3. **Name and address of Sponsor:**

DETAILS OF MREC:

4. **Name and address of MREC:**

MREC West Midlands
Birmingham and The Black Country Health Authority
27 Highfield Road
Edgbaston
Birmingham
B15 3DP

5. **MREC Reference Number:**

MREC 02/7/64

Listed below is a complete record of the review undertaken by MREC with the decisions made, dates of decisions and the requirements at each stage of the review:

The Application was reviewed at the meeting on 24th July 2002. A discussion took place and the application was approved subject to review by the Chairman of satisfactory responses to the following:

General Comments

a) Committee noted that the hospital consultant will not be informed that their patient will be involved in the study and as some patients will be under shared care, it was felt that the consultant should be informed out of courtesy.

b) It was noted that there is no sample size calculation. It appeared that an unusually large number of subjects were to be studied and MREC wondered whether this was really necessary.

A: 0264resp doc

A satisfactory response dated 5th August 2002 has now been received from the principal researcher and the application has now been **APPROVED**.

THE FINAL DOCUMENTS AND ARRANGEMENTS APPROVED BY THE MREC

The following items have been approved by the MREC West Midlands:

Letter to inform the Hospital Consultant, dated 5 August 2002
Professional Indemnity, dated 31 July 2002
Application Form, dated 2 July 2002
Referee's Letter, dated 2 July 2002
Diabetes Service Questionnaire, including Patient Information Sheet, final version, dated 2 July 2002
Research Protocol, dated 2 July 2002
Framework for Performance Assessment, undated
Principal Researcher's CV, undated
Methods of initial Recruitment to study
Compensation Arrangements for subjects

A:\0264resp.doc

Date of Approval: 22nd August 2002

Signature of Chairman: Nagaraj Rao Date: 16·09·02

Name: Dr Jammi N Rao, FRCP, FFPHM

A:\0264resp.doc

Appendix D Copy of the Practices Profile Questionnaire

Practice Characteristics

1. What is the total number of patients registered with your practice? ______________
2. Is your practice: Inner city ☐ Suburban ☐ Rural ☐ Mixed ☐
3. Is your practice a training practice? Yes / No
4. Is your practice a PMS practice? Yes / No
5. Does your practice have a lead person for diabetes? Yes / No
6. If yes, what is their name and profession? ______________________________
7. Please fill this table regarding the number of professionals employed by your practice:

Professional	Number of people	WTE hours per week (if possible)
GPs – full time		
GPs – part time		
GP trainees		
Practice nurses		
Nurse practitioners		
Practice manager		
Administration/ Reception staff		
Other – *please specify*		

8. Please fill this table regarding the professionals that regularly attend your practice:

Professional	Number of people	WTE hours per week (if possible)
District Nurses		
Health visitors		
Community Diabetes Specialist Nurse		
Community Dietician		
Community Chiropodist		
Community Physiotherapist		
Community Optometrist		
Community Physiotherapist		
Psychologist/ Counsellor		
Visiting members of a hospital based team		
Other – *please specify*		

Diabetes Care in your practice

9. Are there any GPs in your practice with a special interest in diabetes? Yes / No

10. Are there any nurses in your practice with a special interest in diabetes? Yes / No

11. Does your practice have specific clinics for seeing patients with diabetes? Yes / No

If the answer is No, please go to question 15

12. How often are these clinics held (please state the frequency) ____________

13. Are these sessions: GP led Nurse led Combined Other (*please specify*) ______

14. Does your practice hold any evening or out of hours diabetes clinic? Yes / No

Please go to question 16

15. If there are no specific practice diabetes clinics, when are patients with diabetes seen in your practice?

    Patient initiated appointments specifically for diabetes ☐

    Doctor or nurse initiated appointments specifically for diabetes ☐

    Opportunistic appointments, when the patient is being seen for something other than diabetes ☐

    Other – *please specify* ____________________

16. If a patient does not attend an appointment to the practice, what do you do? *Please tick the appropriate boxes.*

	Diabetes clinic appointment	Routine surgery appointment
Do nothing		
Send another appointment invitation by post		
Contact the patient		
Other – *Please specify*		

Telephone Help line for diabetes

17. Does your practice have a dedicated diabetes telephone help/advice line that patients can ring when they need advice? Yes / No *If the answer is No, please go to question 20*

18. Who supervises the dedicated telephone help line? ____________________

19. When is the dedicated telephone help line available?

Weekdays office hours ☐ Weekdays evenings ☐ Weekend office hours ☐ 24hours/ 7days ☐

Please go to question 22

20. If there is no dedicated telephone help line, when can patients with diabetes contact the practice?

Weekdays office hours ☐ Weekdays evenings ☐ Weekend office hours ☐ 24hours/ 7days ☐

21. Who are patients able to speak to when they contact the practice? ____________________

Screening and diagnosis

22. Does your practice have a policy for screening patients for diabetes? Yes / No

Referrals to specialists

23. Please estimate the proportion of diabetes care that is undertaken by the practice when compared to the referrals made to specialists ____________________

Diabetes annual reviews

24. Please indicate the percentage of diabetic patients that regularly attend the surgery for diabetes reviews; those that attend the hospital and those under shared care:

	Percentage of diabetic patients
Attend hospital diabetes clinic only	
Attend your practice only	
Attend BOTH the hospital and your practice for regular diabetes reviews	

Routine Reviews

25. Apart from annual reviews, does your practice **routinely review** diabetic patients? Yes / No.

If the answer is No, please go to question 28.

26. If your practice routinely reviews diabetic patients, how often is it done? ________________

27. Who does these routine reviews? ____

28. How long do the consultations for diabetic patients last on average? *Please write the average duration in the appropriate boxes.*

	Average duration of a consultation with the GP	**Average duration of a consultation with the nurse**
Newly diagnosed patient		
Routine consultation		
Annual review		

29. In terms of the nurses who consult diabetic patients in your practice, what is the cost of one hour of nurse consultation time?

30. Are diabetic patients asked to have their blood taken for HbA1c a few days before they attend their routine/ annual review? Yes / No

31. Are patients with diabetes provided with glucose self-monitoring diaries? Yes / No

32. What education materials are available for diabetic patients?

____

____

____

Training

33. Are clinical guidelines available for the management of people with diabetes in your practice?

Yes / No

34. Are these guidelines: National / Regional / Local / Primary care only / Applied in hospital

35. Please describe the type of diabetes training that the GPs in your practice have received in the past 2 years.

____

____

____

36. Please describe the type of diabetes training that the nurses in your practice have received in the past 2 years.

____

____

____

Thank you for your help.

Once completed, please use the freepost envelop provided and post this questionnaire.

Appendix E Standards of Diabetes Care

(Source: The Department of Health 2002: 5)

Standards table

Standard	
Standard 1: Prevention of Type 2 diabetes	**1.** The NHS will develop, implement and monitor strategies to reduce the risk of developing Type 2 diabetes in the population as a whole and to reduce the inequalities in the risk of developing Type 2 diabetes.
Standard 2: Identification of people with diabetes	**2.** The NHS will develop, implement and monitor strategies to identify people who do not know they have diabetes.
Standard 3: Empowering people with diabetes	**3.** All children, young people and adults with diabetes will receive a service which encourages partnership in decision-making, supports them in managing their diabetes and helps them to adopt and maintain a healthy lifestyle. This will be reflected in an agreed and shared care plan in an appropriate format and language. Where appropriate, parents and carers should be fully engaged in this process.
Standard 4: Clinical care of adults with diabetes	**4.** All adults with diabetes will receive high-quality care throughout their lifetime, including support to optimise the control of their blood glucose, blood pressure and other risk factors for developing the complications of diabetes.
Standards 5 & 6: Clinical care of children and young people with diabetes	**5.** All children and young people with diabetes will receive consistently high-quality care and they, with their families and others involved in their day-to-day care, will be supported to optimise the control of their blood glucose and their physical, psychological, intellectual, educational and social development. **6.** All young people with diabetes will experience a smooth transition of care from paediatric diabetes services to adult diabetes services, whether hospital or community-based, either directly or via a young people's clinic. The transition will be organised in partnership with each individual and at an age appropriate to and agreed with them.
Standard 7: Management of diabetic emergencies	**7.** The NHS will develop, implement and monitor agreed protocols for rapid and effective treatment of diabetic emergencies by appropriately trained health care professionals. Protocols will include the management of acute complications and procedures to minimise the risk of recurrence.
Standard 8: Care of people with diabetes during admission to hospital	**8.** All children, young people and adults with diabetes admitted to hospital, for whatever reason, will receive effective care of their diabetes. Wherever possible, they will continue to be involved in decisions concerning the management of their diabetes.
Standard 9: Diabetes and pregnancy	**9.** The NHS will develop, implement and monitor policies that seek to empower and support women with pre-existing diabetes and those who develop diabetes during pregnancy to optimise the outcomes of their pregnancy.
Standards 10, 11 & 12: Detection and management of long-term complications	**10.** All young people and adults with diabetes will receive regular surveillance for the long-term complications of diabetes. **11.** The NHS will develop, implement and monitor agreed protocols and systems of care to ensure that all people who develop long-term complications of diabetes receive timely, appropriate and effective investigation and treatment to reduce their risk of disability and premature death. **12.** All people with diabetes requiring multi-agency support will receive integrated health and social care.

2?3

Blank
In
Original

Appendix F Data for the 14 surgeries used in the main analysis

Table F.1: Data on Outputs and Outcomes (Part I)

Surgeries	(A) Number of Diabetics	(B) Diabetics aged 75 years old or more	(C) Diabetics diagnosed less than 1 year ago	(D) Diabetics with complications	(E) Diabetics with Annual Review	(F) Diabetics with Ischaemic Heart Disease (IHD)	(G) Diabetics with IHD on Statins	(H) Diabetics aged 30-75, without IHD with latest cholesterol >5	(I) Diabetics aged 30-75, without IHD, With cholesterol >5, on Statins	(K) Patients on targeted Statins (G+I)
1	426	107	48	205	307	90	49	70	11	60
2	346	74	49	186	208	66	41	43	20	61
3	211	33	38	98	137	33	16	44	12	28
4	286	73	54	166	204	43	28	39	11	39
5	214	51	33	116	98	41	23	61	14	37
6	461	83	85	294	277	100	51	110	36	87
7	145	35	25	70	87	18	9	7	3	12
8	114	30	18	52	67	16	9	16	6	15
9	239	59	23	131	189	38	16	77	13	29
11	25	5	7	16	8	9	4	3	2	6
12	51	6	9	17	32	2	0	21	4	4
13	88	20	34	41	44	12	6	9	0	6
23	471	90	59	279	392	111	74	141	40	114
24	972	263	99	522	651	158	84	225	24	108

Table F.2: Data on Outputs and Outcomes (Part II)

Surgeries	(L) Diabetics, aged 40 years old or more	(M) Diabetics on targeted ACEI (Diabetics aged 40 years old or more on ACEI)	(N) Diabetics with latest BP ≤ 140/80 in 1 year	(O) Diabetics with latest Cholesterol ≤ 5 in 1 year	(P) Diabetics with latest HbA1c ≤ 7 in 1 year
1	379	142	96	54	35
2	312	106	148	42	17
3	183	63	68	9	8
4	264	98	171	143	111
5	197	64	81	69	60
6	413	180	156	225	153
7	131	45	36	13	17
8	100	32	9	4	2
9	218	86	159	87	77
11	24	9	6	1	1
12	44	11	34	20	11
13	74	26	9	3	2
23	416	185	270	216	161
24	840	356	518	328	352

Table F.3: Data for the GP and Nurse Consultations

Surgeries	Number of Nurse Consultations at surgery	Number of GP consultations at surgery	Number of Home visits by Nurse	Number of Home visits by GP	Number of telephone consultations by Nurse	Number of telephone consultations by GP	Total WTE GP (in hours)	Total WTE Nurse (in hours)
1	1587.00	2715.00	0.00	438.00	118.00	36.00	788.51	503.57
2	2260.00	1941.00	2.00	217.00	185.00	578.00	613.15	720.95
3	1420.00	1429.00	13.00	73.00	104.00	186.00	367.24	455.95
4	1539.00	1706.00	22.00	129.00	55.00	298.00	472.12	487.53
5	1132.00	1169.00	0.00	55.00	0.00	61.00	282.21	344.51
6	1669.00	2367.00	0.00	166.00	1045.00	59.00	586.85	690.28
7	755.00	937.00	33.00	63.00	15.00	113.00	246.63	246.74
8	681.00	780.00	1.00	0.00	0.00	0.00	163.28	207.69
9	906.00	1126.00	29.00	235.00	29.00	84.00	365.13	293.40
11	72.00	266.00	0.00	0.00	1.00	1.00	55.86	22.09
12	208.00	438.00	2.00	0.00	2.00	0.00	91.69	64.52
13	461.00	520.00	0.00	1.00	0.00	2.00	109.69	140.30
23	1305.00	3424.00	4.00	174.00	88.00	468.00	883.39	414.25
24	6371.00	6137.00	8.00	539.00	264.00	777.00	1683.48	1988.45

Table F.4: Data for the Number of Medication Items Prescribed

Surgeries	Number of items of Biguanides prescribed	Number of items of Sulphonylureas prescribed	Number of other anti-diabetic drug items prescribed	Number of items of screening and monitoring prescribed	Number of items for treatment of hypoglycaemia prescribed	Number of items of short-acting insulin prescribed	Number of items of intermediate & long-acting insulin prescribed	Number of items of ACEI prescribed	Number of items of statins prescribed
1	2025	1275	297	1301	61	115	605	1924	1482
2	1517	699	206	1184	42	158	470	1456	1976
3	826	671	161	750	56	112	384	832	572
4	1356	475	490	957	27	178	468	1300	1495
5	976	870	94	800	25	95	387	845	572
6	2231	1611	229	1938	85	238	702	2431	1976
7	630	485	307	706	31	154	267	637	390
8	434	407	60	397	18	32	181	429	442
9	922	1009	123	731	21	70	367	1157	507
11	136	161.25	67.08	121.77	15	22.62	34.76	117	104
12	204	213.75	61.92	247.23	15	16.38	123.24	143	104
13	278	207	114	274	2	38	127	351	143
23	1207	1336	210	1377	52	76	974	2405	1521
24	3867.231	4131.4	351.86	5076.4173	210.163	1702.34	2417.7	4680	1872

Table F.5: Unit Cost of Inputs

Surgeries	Unit cost of WTE Nurse (in hours)	Unit cost of WTE GP (in hours)	Average unit cost of Biguanides (in £)	Average unit cost of Sulphonylureas (in £)	Average unit cost of other anti-diabetic drugs (in £)	Average unit cost of screening and monitoring items (in £)	Average unit cost of hypoglycaemic items (in £)	Average cost of short-acting insulin (in £)	Average unit cost of intermediate& Long-acting insulin (in £)	Average unit cost of ACEI (in £)	Average unit cost of Statins (in £)	Weighted Average unit cost of all items of medication (in £)
1	23.00	86.00	2.60	7.44	27.40	14.69	16.41	53.52	33.48	9.70	23.86	13.59
2	23.00	86.00	2.14	3.99	24.13	14.07	4.84	32.81	35.96	9.70	23.86	14.44
3	23.00	86.00	2.43	5.60	22.05	15.52	4.38	46.24	32.01	9.70	23.86	13.84
4	23.00	86.00	2.06	5.04	29.56	16.08	15.60	31.84	31.09	9.70	23.86	15.41
5	23.00	86.00	2.46	6.87	30.01	16.45	5.91	38.24	35.87	9.70	23.86	13.69
6	23.00	86.00	2.04	4.40	31.96	12.58	4.18	28.20	30.19	9.70	23.86	12.45
7	23.00	86.00	2.12	5.64	26.49	14.85	5.69	29.63	28.05	9.70	23.86	13.98
8	23.00	86.00	2.10	4.73	23.98	16.33	4.59	35.44	38.99	9.70	23.86	14.06
9	23.00	86.00	2.94	7.39	43.70	16.75	3.44	44.50	39.66	9.70	23.86	14.02
11	23.00	86.00	2.00	5.24	26.10	13.96	11.45	56.81	25.59	9.70	23.86	13.50
12	23.00	86.00	1.84	6.18	19.65	15.30	24.34	50.16	37.89	9.70	23.86	14.55
13	23.00	86.00	2.71	4.81	18.34	24.52	2.37	41.95	43.05	9.70	23.86	15.93
23	23.00	86.00	2.02	6.48	28.84	18.58	5.38	33.76	36.25	9.70	23.86	15.35
24	23.00	86.00	2.24	5.09	23.67	16.30	8.90	34.64	33.36	9.70	23.86	14.49

Data estimated using the answers from the Diabetes Services Questionnaire

Table F.6: Sample size and response rate

Surgeries	Diabetics	Minimum sample size at 95% confidence	Actual sample size used	Questionnaires received	Response rate
1	426	202	203	131	64.53%
2	346	182	183	118	64.48%
3	211	136	137	86	62.77%
4	286	164	165	153	92.73%
5	214	138	138	103	74.64%
6	461	210	210	149	70.95%
7	145	105	106	75	70.75%
8	114	88	114	52	45.61%
9	239	148	148	109	73.65%
11	25	24	25	14	56.00%
12	51	45	51	22	43.14%
13	88	72	88	37	42.05%
23	471	212	212	120	56.60%
24	972	276	400	218	54.50%
Total	**4049**	**2002**	**2180**	**1387**	**62.31%**

Table F.7: Patients who have had all the essential services in the previous 12 months

Surgeries	Estimated proportion of diabetics who have had all the essential services in the previous 12 months	95% confidence interval Lower limit	95% confidence interval Upper limit
1	84.73%	79.60%	89.86%
2	88.98%	84.39%	93.58%
3	84.88%	79.04%	90.72%
4	90.20%	86.98%	93.41%
5	78.64%	72.93%	84.35%
6	80.54%	75.30%	85.77%
7	72.00%	64.92%	79.08%
8	75.00%	66.28%	83.72%
9	71.56%	65.30%	77.82%
11	64.29%	47.29%	81.28%
12	59.09%	43.44%	74.74%
13	78.38%	68.22%	88.53%
23	77.50%	71.04%	83.96%
24	67.89%	62.43%	73.35%

Table F.8: Patient understanding about diabetes and its management

Surgeries	Average estimated proportion of patients who understand enough about diabetes and its management	95% confidence interval Lower limit	95% confidence interval Upper limit
1	77.82%	71.48%	84.16%
2	80.98%	74.86%	87.10%
3	80.66%	73.65%	87.67%
4	80.52%	75.72%	85.32%
5	75.60%	68.72%	82.48%
6	78.24%	72.34%	84.14%
7	82.80%	76.40%	89.20%
8	76.98%	67.84%	86.12%
9	74.32%	67.65%	80.99%
11	84.55%	69.76%	99.33%
12	76.63%	62.29%	90.97%
13	76.90%	65.24%	88.55%
23	76.00%	68.80%	83.21%
24	77.44%	72.17%	82.71%

Table F.9: Quality of the diabetes services provided at the surgery

Surgeries	Average estimated proportion of patients who answered 'Always' or 'Most times' regarding the quality of diabetes services provided at the surgery	95% confidence interval Lower limit	95% confidence interval Upper limit
1	83.99%	78.54%	89.43%
2	84.20%	78.61%	89.78%
3	79.15%	72.10%	86.20%
4	90.47%	87.10%	93.85%
5	85.97%	80.77%	91.17%
6	76.32%	70.43%	82.21%
7	82.75%	76.36%	89.14%
8	78.46%	69.89%	87.04%
9	77.74%	71.56%	83.93%
11	82.74%	68.12%	97.35%
12	88.06%	77.15%	98.97%
13	74.71%	63.05%	86.37%
23	80.41%	74.05%	86.76%
24	86.10%	81.65%	90.55%

Table F.10: Patient Empowerment regarding diabetes control

Surgeries	Average estimated proportion of patients who say that they feel 'Confident' or 'Very Confident' in taking good control of their diabetes	95% confidence interval Lower limit	95% confidence interval Upper limit
1	87.60%	82.84%	92.35%
2	90.60%	86.29%	94.91%
3	94.19%	90.37%	98.00%
4	94.70%	92.24%	97.16%
5	92.00%	88.11%	95.89%
6	89.73%	85.65%	93.80%
7	88.00%	82.87%	93.13%
8	88.46%	82.03%	94.89%
9	89.72%	85.43%	94.01%
11	100.00%	100.00%	100.00%
12	81.82%	69.54%	94.09%
13	94.44%	88.66%	100.23%
23	86.55%	81.25%	91.86%
24	90.23%	86.73%	93.74%

Table F.11: Consultations with Diabetes-related Specialists

Surgeries	Average Number of times each diabetic has had a consultation with a Ophthalmologist in the previous 12 months	Average Number of times each diabetic has had a consultation with a Diabetologist in the previous 12 months	Average Number of times each diabetic has had a consultation with a Dietician in the previous 12 months
1	1.73	0.2	0.96
2	1.52	0.14	0.92
3	1.73	0.17	1.19
4	1.52	0.4	1.83
5	1.49	0.38	0.91
6	1.48	0.12	0.68
7	1.28	0.37	1.03
8	1.43	0.17	0.99
9	1.43	0.34	0.77
11	1.5	1	1
12	1.18	0.41	1.82
13	1.27	0.14	0.7
23	1.35	0.44	0.71
24	1.88	0.15	0.85

Appendix G Data for the additional 9 surgeries used for robustness analysis

Table G.1: Data on Outputs and Outcomes (Part I)

Surgeries	(A) Number of Diabetics	(B) Diabetics aged 75 years old or more	(C) Diabetics diagnosed less than 1 year ago	(D) Diabetics with complications	(E) Diabetics with Annual Review	(F) Diabetics with Ischaemic Heart Disease (IHD)	(G) Diabetics with IHD on Statins	(H) Diabetics aged 30-75, without IHD with latest cholesterol >5	(I) Diabetics aged 30-75, without IHD, With cholesterol >5, on Statins	(K) Patients on targeted Statins (G+I)
14	337	59	63	189	224	69	41	30	9	50
15	217	41	46	117	11	33	25	41	3	28
16	260	57	39	146	64	50	34	3	0	34
17	202	45	39	113	1	42	22	28	8	30
18	83	14	10	59	54	23	13	4	8	21
19	175	47	26	67	102	67	20	29	5	25
20	324	58	74	161	225	56	38	6	0	38
21	48	13	44	33	24	20	12	3	0	12
22	88	10	19	39	71	11	6	10	3	9

Table G2: Data on Outputs and Outcomes (Part II)

Surgeries	(L) Diabetics, aged 40 years old or more	(M) Diabetics on targeted ACEI (Diabetics aged 40 years old or more on ACEI)	(N) Diabetics with latest BP ≤ 140/80 in 1 year	(O) Diabetics with latest Cholesterol ≤ 5 in 1 year	(P) Diabetics with latest HbA1c ≤ 7 in 1 year
14	308	118	62	10	3
15	178	78	76	87	33
16	234	111	21	131	75
17	189	77	47	113	93
18	78	27	17	3	2
19	162	47	48	92	77
20	287	95	16	4	9
21	47	17	10	5	8
22	61	25	10	6	6

Table G.3: Data for the GP and Nurse Consultations

Surgeries	Number of Nurse Consultations at surgery	Number of GP consultations at surgery	Number of Home visits by Nurse	Number of Home visits by GP	Number of telephone consultations by Nurse	Number of telephone consultations by GP	Total WTE GP (in hours)	Total WTE Nurse (in hours)
14	2012	5963	Not available	Not available	Not available	Not available	1248.255	612.3187
15	1204	1208	Not available	Not available	Not available	Not available	252.8747	366.4173
16	1711	3346	Not available	Not available	Not available	Not available	700.4293	520.7143
17	1504	2692	Not available	Not available	Not available	Not available	563.5253	457.7173
18	380	529	Not available	Not available	Not available	Not available	110.7373	115.6467
19	612	1050	Not available	Not available	Not available	Not available	219.8	186.252
20	1520	1944	Not available	Not available	Not available	Not available	406.944	462.5867
21	3	78	Not available	Not available	Not available	Not available	16.328	0.913
22	587	1139	Not available	Not available	Not available	Not available	238.4307	178.6437

Table G.4: Data for the Number of Medication Items Prescribed

Surgeries	Number of items of Biguanides prescribed	Number of items of Sulphonylureas prescribed	Number of other anti-diabetic drug items prescribed	Number of items of screening and monitoring prescribed	Number of items for treatment of hypoglycaemia prescribed	Number of items of short-acting insulin prescribed	Number of items of intermediate & long-acting insulin prescribed	Number of items of ACEI prescribed	Number of items of statins prescribed
14	1135	613	53	1164	30	164	656	1547	1534
15	756	641	10	1100	19	244	635	1079	819
16	1321	962	42	1025	29	151	528	1482	1300
17	964	892	49	731	22	104	376	1014	832
18	415	186	38	259	8	86	237	351	403
19	599	476	17	582	31	115	319	611	520
20	1088	848	280	1249	57	202	628	1274	962
21	246	105	4	249	2	35	132	221	208
22	226	164	18	254	12	83	182	364	299

Table G.5: Unit Cost of Inputs

Surgeries	Unit cost of WTE Nurse (in £)	Unit cost of WTE GP (in £)	Average unit cost of Biguanides (in £)	Average unit cost of Sulphonylureas (in £)	Average unit cost of other anti-diabetic drugs (in £)	Average unit cost of screening and monitoring items (in £)	Average unit cost of hypoglycaemic items (in £)	Average cost of short-acting insulin (in £)	Average unit cost of intermediate& Long-acting insulin (in £)	Average unit cost of ACEI (in £)	Average unit cost of Statins (in £)
14	23.00	86.00	2.99	5.86	26.79	18.06	7.48	37.4	38.46	9.7	23.86
15	23.00	86.00	2.01	3.45	37.8	17.75	8.49	25.43	25.88	9.7	23.86
16	23.00	86.00	1.87	4.37	16.2	16.01	14.59	36.58	37.36	9.7	23.86
17	23.00	86.00	1.69	4.4	29.63	16.24	6.39	26.87	25.9	9.7	23.86
18	23.00	86.00	1.99	2.96	20.09	16.16	7.05	28.94	27.99	9.7	23.86
19	23.00	86.00	2.76	4.84	13.86	20.57	10.89	35.24	29.87	9.7	23.86
20	23.00	86.00	2.1	4.32	17.56	17.46	7.85	36.16	38.85	9.7	23.86
21	23.00	86.00	2.17	3.03	8.75	16.57	17.99	37.32	41.18	9.7	23.86
22	23.00	86.00	2.08	4.28	14.43	18.25	6.44	29.8	26.62	9.7	23.86

Data estimated using the answers from the Diabetes Services Questionnaire

Table G.6: Sample size and response rate

Surgeries	Diabetics	Minimum sample size at 95% confidence	Actual sample size used	Questionnaires received	Response rate
14	337	180	180	100	55.56%
15	217	139	139	110	79.14%
16	260	155	156	97	62.18%
17	202	133	133	118	88.72%
18	83	68	83	42	50.60%
19	175	120	121	106	87.60%
20	324	176	177	105	59.32%
21	48	43	48	22	45.83%
22	88	72	88	30	34.09%
Total	**1734**	**1086**	**1125**	**730**	**64.89%**

Table G.7: Patients who have had all the essential services in the previous 12 months

Surgeries	Estimated proportion of diabetics who have had all the essential services in the previous 12 months
14	74.29%
15	59.09%
16	92.86%
17	86.36%
18	88.00%
19	76.67%
20	85.57%
21	86.44%
22	75.47%

Table G.8: Patient understanding

Surgeries	Average estimated proportion of patients who understand enough about diabetes and its management
14	78.19%
15	80.97%
16	80.39%
17	78.40%
18	83.53%
19	82.79%
20	84.98%
21	71.98%
22	78.78%

Table G.9: Quality of the diabetes services provided at the surgery

Surgeries	Average estimated proportion of patients who answered 'Always' or 'Most times' regarding the quality of diabetes services provided at the surgery
14	80.06%
15	81.04%
16	90.07%
17	89.51%
18	84.42%
19	85.34%
20	80.43%
21	58.30%
22	77.45%

Table G.10: Patient Empowerment regarding diabetes control

Surgeries	Average estimated proportion of patients who say that they feel 'Confident' or 'Very Confident' in taking good control of their diabetes
14	91.00%
15	92.59%
16	92.63%
17	93.16%
18	95.12%
19	95.24%
20	89.32%
21	81.82%
22	90.00%

Table G.11: Consultations with Diabetes-related Specialists

Surgeries	Average Number of times each diabetic has had a consultation with a Ophthalmologist in the previous 12 months	Average Number of times each diabetic has had a consultation with a Diabetologist in the previous 12 months	Average Number of times each diabetic has had a consultation with a Dietician in the previous 12 months
14	1.52	0.42	0.89
15	1.2	0.65	1.28
16	1.46	0.37	0.71
17	1.53	0.2	0.59
18	1.58	0.57	0.66
19	1.49	0.23	0.82
20	1.39	0.45	0.8
21	1.32	1.68	0.77
22	1.1	0.4	0.8

References

Albrecht, G. L. (1994). 'Subjective health assessment'. In C. Jenkinson (eds) 'Measuring health and medical outcomes', London: UCL Press, pp. 7-26.

Acheson Report (1998). 'Independent Inquiry into Inequalities and Health'. The Stationery Office.

Aiken, L. (1997), 'Questionnaires and Inventories: surveying opinions and assessing personality'. New York: John Wiley & Sons, Inc.

Allen, R. Athanassopoulos, A., Dyson, R. G., Thanassoulis, E. (1997). 'Weight restrictions and value judgements in data envelopment analysis: evolution, development and future directions'. *Annals of Operations Research*, **73**: 13-34.

American Diabetes Association (2000). 'Type 2 diabetes in children and adolescents'. *Diabetes Care*, **23**: 381-389.

Baker, R. (1992). 'General practice in Gloucestershire, Avon and Somerset: explaining variations in standards'. *British Journal of General Practice*, **42**: 415-418.

Banker, R. D., Charnes, A, Cooper, W. W. (1984). 'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis'. *Management Science*, **30**: 1078-1092.

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.

Bloor, K., Maynard, A. and Street, A. (2000). 'The cornerstone of Labour's 'new NHS': reforming primary care', in P. Smith (Ed), 'Reforming Markets in Health Care: An Economic Perspective', Buckingham: Open University Press.

Blalock, A. (1999). 'Evaluation and the Performance Management Movement: From Estrangement to Useful Integration?'. *Evaluation* **5** (2): 117-149.

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.

Bodily, S. E. (1978). 'Police Sector Design Incorporating Preferences of Interest Groups for Equality and Efficiency', *Management Science*, **24** (12): 1301-1313.

Bowling, A. (1991). 'Measuring Health: A Review of Quality of Life Measurement Scales'. Buckingham: Open University Press.

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.

British Diabetic Association (2001). 'What diabetes care to expect'. Available online on http://www.diabetes.org.uk/manage/what.htm (Accessed on 22 January 2003).

British Medical Association (2003). 'Investing in General Practice: The New General Medical Services Contract'.

Bryce, C., Engberg, J. and Wholey, D. (2000). 'Comparing the Agreement Among Alternative Models in Evaluating HMO Efficiency'. *Health Services Research* **35**: 509-528.

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, pp. 129-144.

Campbell, S.M., Roland, M.O., Quayle, J.A., Buetow, S.A. and Shekelle, P.G. (1998). 'Quality indicators for general practice: which ones can general practitioners and health authority managers agree are important and how useful are they?' *Journal of Public Health Medicine* **20**: 414-421.

Carter, N., Klein, R. and Day, P. (1992). 'How Organisations Measure Success: The Use of Performance Indicators in Government'. London: Routledge.

Casalino, L. P. (1999). 'The unintended consequences of measuring quality on the quality of medical care'. *New England Journal of Medicine*, **341** (15): 1147-1150.

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

Charnes, A., Cooper W., Huang, Z. and Sun, D. (1990). 'Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of Econometrics,* **46**: 73-91.

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.

Connolly, V. and Kesson, C. (1996). 'Socioeconomic status and clustering of cardiovascular disease risk factors in diabetic patients'. *Diabetes Care* 1996; **19**: 419-422.

Cook, T. and Reichardt, C. (1979). 'Qualitative and Quantitative Methods in Evaluation Research'. Beverly Hills, CA: Sage Publications.

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.

Crombie, I. K. and Davies, H. T. O. (1998). 'Beyond health outcomes: the advantages of measuring process'. *Journal of Evaluation in Clinical Practice,* **4** (1): 31-38.

Culyer, A. J. (1976). 'Need and the National Heath 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.

Davies, H. T. O. and Crombie, I. K. (1997). 'Interpreting health outcomes'. *Journal of Evaluation in Clinical Practice*, **3** (3): 187-199.

Davies, H. T. O. and Mannion, R. (2000). 'Clinical Governance: Striking a Balance between Checking and Trusting'. In P.C. Smith (eds). 'Reforming markets in health care: an economic perspective'. Buckingham: Open University Press, pp. 247-267.

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.

Donabedian, A. (1980). 'The Definition of Quality and Approaches to its Assessment'. Ann Arbor, Michigan: Health Administration Press.

Dowell, T. and Neal, R. (2000). 'Vision and change in primary care: past, present and future', in P. Tovey (Ed.), 'Contemporary Primary Care: the challenges of change'. Buckingham: Open University Press.

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.

Dyson, R. G. and Thanassoulis, E. (1988). 'Reducing weight flexibility in data envelopment analysis'. *Journal of the Operational Research Society*, **39**: 563-576.

Dyson, R. G., Allen, R., Camanho, A., Podinovski, V., Sarrico, C. and Shale, E. (2001). 'Pitfalls and protocols in DEA'. *European Journal of Operational Research*, **132** (2): 245-259.

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.

Fernandez-Castro, A. and Smith, P.C. (1994). 'Towards a general non-parametric model of corporate performance'. *Omega: International Journal of Management Science*, **22**: 237-249.

Fisher, C. (1998). 'Resource Allocation in the Public Sector: Values, priorities and markets in the management of public services'. London: Routledge.

Fitzgerald, L. and Storbeck, J. E. (2002). 'Distinguishing interests in the performance of regulated water: the UK experience'. In A. Neeley, A. Walters and R. Austin (eds). 'Performance Measurement Research and Action'. Centre for Business Performance, Cranfield, UK, pp. 197-203.

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.

Fry, J. (1993). General Practice – The Facts, Oxford: Radcliffe Medical Press.

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.

Gardner, D. (1998). 'Performance measurement' in J. Wilson (Ed.), 'Financial Management for the Public Sector'. Buckingham: Open University Press.

Gilthorpe, M. S. and Wilson, R. C. (2003). 'Rural/ urban differences in the association between deprivation and healthcare utilisation'. *Social Science & Medicine*, **57**: 2055-2063.

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.

Glasgow, R. and Strycker, L. (2000). 'Preventive Care Practices for Diabetes Management in Two Primary Care Samples'. *American Journal of Preventive Medicine*, **19** (1): 9-14.

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. pp. 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.

Golany, B. and Tamir, E. (1996). 'Evaluating Efficiency-Effectiveness-Equality Trade-offs: A Data Envelopment Analysis Approach'. *Management Science*, **41** (7): 1172-1184.

Göni, S. (1999). 'An analysis of the effectiveness of Spanish primary health care teams'. *Health Policy*, **48**: 107-117.

Goyder, E. C., McNally, P. G., Drucquer, M., Spiers, N. and Botha, J. (1998). 'Shifting of care for diabetes from secondary to primary care, 1990-1995: review of general practices. *British Medical Journal*, **316**: 1505-1506.

Greenhalgh, T. and Eversley, J. (1999). 'Quality in General Practice'. London: King's Fund.

Griffin, S. (1998). 'Diabetes care in general practice: meta-analysis of randomised control trials'. *British Medical Journal*, **317**: 390-396.

Grosskopf, S. and Valdmanis, V. (1987). 'Measuring hospital performance: A non-parametric approach'. *Journal of Health Economics*: **6**: 89-108.

Guba, E. and Lincoln, Y. (1981). 'Effective Evaluation: Improving the Usefulness of Evaluation Results Though Responsive and Naturalistic Approaches'. San Francisco: Jossey-Bass.

Haire-Joshu, D., Glasgow, R. and Tibbs, T. (1999). 'Smoking and Diabetes'. *Diabetes Care*, **22** (11): 1887-1898.

Hansson, L., Zanchetti, A., Carruthers, S. Dahlöf, B., Elmfeldt, D., Julius, S., Ménard, J., Rahn, K. H., Wedel, H. and Westerling, S. (1998). 'Effects of intensive blood pressure lowering and low dose aspirin in patients with hypertension: Principle results of the Hypertension Optimal Treatment (HOT) randomised trial'. *The Lancet,* **351**: 1755-1762.

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.

Heart Protection Study Collaborative Group (2003). MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo controlled trial'. *The Lancet*, **361**: 2005-2016.

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.

Home, P., Coles, J., Goldacre, M., Mason, A. and Wilkinson, E. (1999) (eds). 'Health Outcome Indicators: Diabetes'. Report of a Working Group to the Department of Health. Oxford: National Centre for Health Outcomes Development.

Hood, C. (1995). 'The "New Public Management" in the 1980s: Variations on a Theme'. *Accounting, Organizations and Society* **20** (2/3): 93-109.

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.

Khunti, K., Ganguli, S., Baker, S. and Lowy, A. (2001). 'Features of primary care associated with variations in process and outcome of care in people with diabetes'. *British Journal of General Practice*, **51**: 356-360.

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, pp. 1-6.

Johnson, P. E., Veazie, P. J., Kochevar, L., O'Connor, P. J., Potthoff, S. J., Verma, D. and Dutta, P. (2002). 'Understanding Variation in Chronic Disease Outcomes'. *Health Care Management Science* **5** (3):175-189.

Kelly, W., Mahmood, R., Kelly, M., Turner, S. and Elliott, K. (1993). 'Influence of social deprivation on illness in diabetic patients'. *British Medical Journal*; **307**: 1115-1116.

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.

Le Grand, J., Mays, N., Mulligan, J. (Ed.) (1998). 'Learning from the NHS Internal Market'. London: King's Fund.

Levitt, R. and Wall, A. (1984). 'The Reorganised National Health Service'. 3rd Ed. London: Croom Helm.

Lloyd, C., Dyer, P. and Barnett, A. (2000). 'Prevalence of symptoms of depression and anxiety in a diabetes clinic population'. *Diabetic Medicine*, **17**: 198-202.

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, pp. 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.

McColl, A., Roderick, P., Gabbay, J., Smith, H. and Moore, M. (1998). 'Performance indicators for primary care groups: an evidence based approach'. *British Medical Journal* **317**: 1354-60.

Maital, S. (1973). 'Public Goods and Income Distribution: Some Further Results'. *Econometrica*, 561-568.

Maital, S. (1975). 'Appointment of Public Goods Benefits to Individuals'. *Public Finance*, **3**: 397-416.

Mandell, M. (1991). 'Modelling Effectiveness-Equity Trade-offs in Public Service Delivery Systems'. *Management Science*, **37**: 4, 467-482.

Mitchell, R. Gleave, S., Bartley, M., Wiggins, D. and Joshi, H. (2000). 'Do attitude and area influence health? A multilevel approach to health inequalities. *Health & Place*, **6**: 67-79.

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.

Netten, A., Rees, T. and Harrison, G. (2001). 'Unit Costs of Health and Social Care 2001'. Personal and Social Services Research Unit, University of Kent, Kent.

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.

Okun, A. (1975). 'Equality and Efficiency: The Big Trade-off'. Brookings Institute: Washington.

Olesen, O. and Petersen, N. (1995). 'Chance Constrained Efficiency Evaluation'. *Management Science*, **41**: 442-457.

Orme, C. and Smith, P. C. (1996). 'The Potential for Endogeneity Bias in Data Envelopment Analysis'. *Journal of the Operational Research Society*, **47**: 73-83.

Ovhed, I., Johansson, E., Odeberg, H. and Råstam, L. (2000). 'A Comparison of Two Different Team Models for Treatment of Diabetes Mellitus in Primary Care'. *Scandinavian Journal of Caring Science*, **14**: 253-258.

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.

Patton, M.Q. (1986). 'Utilization-Focused Evaluation'. 2nd ed. Neward Park CA: Sage Publications.

Patrick, D.L. and Erickson, P.E. (1993). 'Health status and health policy: allocating resources to health care. Oxford: Oxford University Press.

Pawson, R. and Tilley, N. (1997), 'Realistic Evaluation'. London: Sage Publications.

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.

Pedraja-Chaparro, F., Salinas-Jiménez, J., Smith, P. C. (1999). 'On the quality of the data envelopment analysis model'. *Journal of the Operational Research Society* **60**: 636-644.

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.

Podinovski, V. (1999). 'Side effects of absolute weight bounds in DEA models'. *European Journal of Operational Research* **66**: 109-138.

Podinovski, V. (2001). 'Validating absolute weight bounds in Data Envelopment Analysis (DEA) models', *Journal of the Operational Research Society* **52**: 221-225.

Podinovski, V. (2002). 'Weight Restrictions and Radial Measures of Efficiency'. Warwick Business School Research Paper No. 352: 1-29.

Podinovski, V. and Athanassopoulos, A. (1998). 'Assessing the relative Efficiency of decision making units using DEA models with weight restrictions'. *Journal of the Operational Research Society* **49**: 500-508.

Pollitt, C. (1986). 'Beyond the managerial model: the case for broadening performance assessment in Government and public services', *Financial Accountability & Management* **2**: 155-170.

Porto, J. V. (2000). 'Comparing the Efficiency and Effectiveness of Substance Abuse Treatment Programs Using a Two-stage Data Envelopment Analysis'. PhD thesis.

Proctor, S. and Campbell, J. (1999). 'A developmental performance framework for primary care'. *International Journal of Health Quality Assurance*, **12** (7): 279-286.

Prescribing Price Authority (2002). 'Cardiovascular Prescribing', Prescribing Analysis and Cost Centre Pages 4th Quarter 2001/2002. Available online on http://www.ppa.org.uk/news/pact-012002.htm (Accessed on 21 January 2003).

Reichard, P (1992). 'Risk factors for progression of microvascular complications in the Stockholm Diabetes Intervention Study'. *Diabetes Research and Clinical Practice*, **16**: 151-156.

Roper, N., Bilous, R., Kelly, W., Unwin, C. and Connolly, V. (2001). 'Excess mortality in a population with diabetes and the impact of material deprivation: longitudinal, population based study'. *British Medical Journal*, **322**: 1389-1393.

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-Jimenez, 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, pp. 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.

Scriven, M. (1967). 'The methodology of evaluation', in R.W. Tyler, R.M. Gagne and M. Scriven (eds), 'Perspectives on Curriculum Evaluation', Chicago: Rand McNally. pp. 39-83.

Secretary of State for Health and Others (1989). 'Working for Patients'. London: HMSO.

Sen, A. (1992). 'Inequality Re-examined'. Oxford: Claredon Press.

Sengupta, J. (1996). 'The Efficiency distribution approach in Data Envelopment Analysis'. *Journal of the Operational Research Society*, **47**: 1387-1397.

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, K. and Sheaff, R. (2000). 'Some Are More Equal Than Others: Differing Levels of Representation in Primary Care Groups'. In A. Hann (Ed.), 'Analysing Health Policy', Aldershot: Ashgate.

Smith, P. C. (1990). 'The Use of Performance Indicators in the Public Sector'. *Journal of the Royal Statistical Society*, **153** (Part I): 53-72.

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. (1993). 'Outcome-related Performance Indicators and Organizational Control in the Public Sector'. *British Journal of Management*, **4**: 135-151.

Smith, P. C. (1995). 'On the unintended consequences of publishing performance data in the public sector'. *International Journal of Public Administration*, **18**: 277-310.

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. pp. 1-19.

Smith, P.C. (1997). 'Model Misspecification in Data Envelopment Analysis'. *Annals of Operations Research*, **73**: 233-252.

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, pp. 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.

St Vincent Working Group (1996). 'St Vincent and improving diabetic care'. *Diabetic Medicine*, **13**: S1-S128.

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.

Thanassoulis, E., Boussofiane, A. and Dyson, R. G. (1996). 'A Comparison of Data Envelopment Analysis and Ratio Analysis as Tools for Performance Assessment', *Omega*, **24** (3): 229-244.

The Audit Commission (2000a). 'On target: the practice of performance indicators'. London.

The Audit Commission (2000b). 'Testing Times: A review of diabetes services in England and Wales'. National Report, London: The Audit Commission.

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).

The Department of Health (2001a). 'National service framework for diabetes standards'. London: Department of Health.

The Department of Health (2001b). 'The Expert Patient: A New approach to Chronic Diseases Management for the 21st Century'. Available online on http://www.Department of Health.gov.uk/healthinequalities/ep_report.pdf (accessed 21 January 2003).

The Department of Health (2003). 'National service framework for diabetes delivery strategy'. London: Department of Health.

The Diabetes Control and Complication Trial Research Group (1993). 'The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine*, **329**: 977-986.

The Diabetes Integrated Care Team (1994). 'Integrated care for diabetes: clinical, psychosocial, and economic evaluation'. *British Medical Journal* **308**: 1208-12.

The Heart Outcome Prevention Evaluation Study Investigators (2000). 'Effects of ramipril on cardiovascular and microvascular outcome in people with diabetes; results from the HOPE study and the MICRO-HOPE sub-study'. *The Lancet*, **355**: 252-259.

The PMS National Evaluation Team (2002). 'National Evaluation of First Wave NHS Personal Medical Services Pilots: Summaries of Findings from Four Research Projects'. Document submitted to the Department of Health.

Thompson, R., Singleton, F., Thrall, R. and Smith, B. (1986). 'Comparing site evaluations for locating a high-energy physic in Texas'. *Interfaces*, **16**: 35-49.

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.

Tudor-Hart, J. (1971). 'The inverse care law'. *The Lancet*, **27**: 405-12.

UK Prospective Diabetes Study Group (UKPDS) (1990). 'Complications in newly diagnosed type 2 diabetic patients and their association with different clinical and biochemical risk factors' (UKPDS 6). *Diabetes Research*, **13**: 1-11.

UK Prospective Diabetes Study Group (UKPDS) (1998a). 'Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with Type 2 diabetes mellitus' (UKPDS 33). *The Lancet*, **352**: 837-853.

UK Prospective Diabetes Study Group (UKPDS) (1998b). 'Tight blood pressure control and risk of macrovascular and microvascular complications in Type 2 diabetes' (UKPDS 38). *British Medical Journal*, **317**: 703-713.

Unwin, N., Binns, D., Elliott, K. and Kelly, W. (1996). 'The relationships between cardiovascular risk factors and socio-economic status in people with diabetes'. *Diabetic Medicine*; **13**: 72-79.

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 utilzan en Atención Primaria?. La opinión de los professionales de 3 áreas sanitarias'. *Atención Primaria* **20** (4): 191-94.

Wake, N., Hisashige, A., Katayama, T., Kishikawa, H., Ohkubo, Y., Sakai, M, Araki, E. and Shichiri, M. (2000). Cost-Effectiveness of intensive insulin therapy for type 2 diabetes: a 10-year follow-up of the Kumamoto study'. *Diabetes Research and Clinical Practice*, **48**: 201-210.

Wareham, N. J. and Griffin, S. J. (2001). Should we screen for type 2 diabetes? Evaluation against National Screening Committee criteria'. *British Medical Journal*, **322**: 986-988.

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. pp. 20-33.

Wong, Y-H.B. and Beasley, J.E. (1990). 'Restricting weight flexibility in data envelopment analysis'. *Journal of the Operational Research Society*, **41**: 829-835.

World Health Organisation (1994). 'Prevention of diabetes mellitus: report of a WHO study group. Technical report series N. 844. Geneva: WHO, 1994.

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.

Wholey, J. (1996). 'Formative versus Summative Evaluation: Related Issues in Performance Measurement'. *Evaluation Practice*, **17** (2): 145-149.

Wilkin, D., Hallam, L., Leavey, R. and Metcalfe, D. (1987). 'Anatomy of Urban General Practice'. London: Tavistock Publications.

Williams, R., Baxter, H., Airey, M., Ali, S. and Trudi, D. (2000). 'Diabetes UK funded surveys of the structural provision of diabetes services in the UK. Document produced by Diabetes UK and Nuffield Institute.

Van den Hombergh, P., Grol, R., van den Hoogen, H.J.M., and van den Bosch, W.J.H.M. (1998). 'Assessment of management in general practice: validation of a practice visit method'. *British Journal of General Practice*, **48**: 1743-50.

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.