Evidence

The Health Foundation
Inspiring Improvement

Commissioning health

A comparison of English primary care trusts
Preliminary statistical analysis

Stephen Martin and Peter C Smith
February 2010

Identify Innovate Demonstrate Encourage

Acknowledgements

This study was undertaken as part of the Quest for Quality and Improved Performance (QQUIP), a five-year initiative of the Health Foundation. We are grateful to our colleagues Rowena Jacobs and Andrew Street, and to Vin McLoughlin, Dale Webb and Sarah Campbell at the Health Foundation, for helpful comments and advice. We are also grateful to Daniel Eayres at NCHOD for assistance with the mortality data and Linda Williams at the Office for National Statistics for providing census data for the new (October 2006) PCT boundaries.

Commissioning health. A comparison of English primary care trusts
Preliminary statistical analysis

Stephen Martin* and Peter C Smith**

*Department of Economics, University of York, Heslington, York YO10 5DD

**Imperial College Business School and Institute for Global Health, Exhibition Road, London SW7 2AZ

Published by:

The Health Foundation
90 Long Acre
London WC2E 9RA
Telephone: 020 7257 8000
Facsimile: 020 7257 8001

www.health.org.uk

Registered charity number 286967
Registered company number 1714937

First published 2010

ISBN 978-1-906461-19-5

Contents

Foreword

As the NHS faces up to the economic crisis and its impact on the public sector, it is increasingly important to know how the service is currently using resources and where efficiency gains can sensibly be made. In considering this, we need to look at the system as a whole as well as its constituent parts. In the English NHS, this means that we must scrutinise the performance not just of healthcare providers but also of commissioners – in this case, the primary care trusts (PCTs) purchasing healthcare services on behalf of their populations.

The Health Foundation has funded Peter Smith, formerly of the Centre for Health Economics at the University of York and now at Imperial College, to undertake economic analysis of healthcare budgeting, as we believe this can be useful in providing information to aid the decision-making process at system level and within PCTs.

There are a number of caveats that come with this type of analysis. Most important, the findings of this report give us a comparison of performance between PCTs, with the intention of indicating how well each PCT is performing relative to its peers. According to the methodology, any PCT which records the largest output per unit of input, for any output, will be rated as 100 per cent efficient, irrespective of its absolute level of performance. This is an econometric analysis of efficiency, not 'efficiency' as used in the everyday sense.

This is preliminary research, which brings together a range of different methods. Using the methods together acts as a check of validity, and this has proven to be the case. The various methods are coming up with consistent results, which is reassuring. The methods are better employed to compare results between organisations over time but as yet adequate longitudinal data are not available. At present, we have a 'snapshot' of information on comparative efficiency and this report is a good starting point.

The report is clear that while, as in any comparative study, there are some differences in efficiency between PCTs, it is unlikely that significant financial efficiencies can be garnered by ridding commissioning of unacceptable variation. Overall, little evidence is found of relative inefficiency in the way that funds are allocated between the major disease groups, as set out in programme budgeting categories. Many governmental efficiency drives are predicated on the notion of wide variations in the performance of public bodies. But this report comes to a different conclusion regarding PCTs: most appear to be performing at a similar level. This suggests that increased efficiency will only come about as a result of step-changes at system level, not through individual PCTs adjusting their internal processes.

The message to regulators and performance managers appears to be that within the current commissioning framework there is precious little in the way of extra efficiency to be found by focusing on individual PCTs. In terms of meeting the challenge of the economic crisis, this report seems to add force to the argument that the approach of cuts across the board will not increase productivity and that system-level change in the way commissioning functions and the incentives it offers for efficiency will need a radically different approach in future.

The report gives us much food for thought. The Health Foundation continues to work with leading international health economists to improve our understanding of the theory and practice of achieving value for money. As researchers are able to look at results over time as more longitudinal data become available, we will be able to explore linkages in the funding system, for example, between investment in prevention and the effect over time on hospital admissions. In addition, we are funding the development of practical tools to assist commissioners in achieving transparent funding decisions with the appropriate involvement of stakeholders.

Much more thought and empirical research is required to help us understand the complex issue of value for money in healthcare and the Health Foundation is proud to be making a contribution to this field.

Martin Marshall
Clinical Director and Director of Research & Development
The Health Foundation

Executive summary

1. The pursuit of efficiency in health services has been a high policy priority for successive governments and is of concern worldwide, as reflected in the World Health Organization's *World Health Report 2000* (WHO, 2000). Efficiency concerns are embedded in the design of the English NHS, for example in the use over several decades of fixed budgets for local health authorities (now primary care trusts (PCTs)), the important gatekeeping role of GPs, and more recent reforms to the hospital payment mechanism (payment by results). Efficiency is now a central concern of the PCT commissioning function.

2. Efficiency has many connotations, but in this report we consider it as equivalent to the notion of 'value for money': that is, the ratio of some valued outputs to the costs expended. Making this apparently straightforward concept operational gives rise to some methodological and practical challenges, but there is a burgeoning academic literature on efficiency measurement, and we deploy some well-established techniques to measure efficiency levels within the NHS.

3. The units of analysis we examine are the 152 PCTs in England. These are responsible, among other things, for commissioning health services and promoting public health for their populations. It is therefore reasonable to presume that the 'valued outputs' they are seeking to achieve are represented in the health of their populations. However, the budgets available to PCTs vary considerably. Furthermore, in pursuing their health objectives, different PCTs may be operating in very different needs environments and therefore face different constraints in advancing population health. Any analysis of efficiency has to recognise and adjust for these factors.

4. The techniques we use seek to measure PCT efficiency while adjusting for uncontrollable constraints on success. They take two broad forms: parametric and non-parametric methods. The approach underlying both methods is to consider a range of measured inputs and valued outputs, and to identify the efficiency of each PCT relative to what seems to be 'best practice' among all the 152 PCTs. In doing so, efforts are made to adjust for the different needs environments in which the PCTs operate. The intention is to indicate how well each PCT is performing relative to its peers. The parametric methods are based on familiar regression techniques. They model PCT costs as a function of inputs and uncontrollable environmental factors, in particular estimates of the level of medical 'needs' in their populations. The assumption is that the deviation between predicted costs and actual costs is an indicator of the efficiency of the PCT. Two parametric techniques are tested: corrected ordinary least squares (COLS), which makes the crude assumption that all variation from the predicted expenditure is due to efficiency variations; and stochastic frontier analysis (SFA), which assumes that some of the variation is due to unmeasured random influences on performance and that some is due to inefficiency.

5. Non-parametric techniques use the same underlying economic model, but employ very different methods. The approach used in this report is known as data envelopment analysis (DEA). For each observation (PCT), DEA uses linear programming methods to search for the other PCT or combination of PCTs that produce the same (or better) outputs and operate in the same (or more challenging) environments, while using lower levels of expenditure.

6. Efficiency measurement techniques have been used extensively in healthcare. However, they have rarely been used to examine the extent to which healthcare organisations succeed in producing health. This study is therefore unusual in seeking to formulate efficiency in terms of the ultimate objective of healthcare, the production of health. We use as inputs the expenditure

incurred by each PCT. As outputs we use mortality indices in two major disease areas: cancer and circulatory disease. We also control for relative clinical needs, using the needs index used by the Department of Health to distribute funds to PCTs.

7. The various parametric models tested yield similar results in the sense that PCTs are ranked similarly whichever parametric method is used. Under COLS, average efficiency is about 88 per cent. However, when account is taken of possible random fluctuation using SFA, the average efficiency increases to between 96 and 97 per cent, depending on what index of mortality is used. Using the SFA model, the estimates of efficiency for individual PCTs range between 87 and 99 per cent.

8. DEA relies heavily on the scores of outlying PCTs and takes no account of possible measurement error. It is therefore not surprising that its efficiency scores indicate a larger range, between 77 and 100 per cent. Furthermore, although DEA scores are strongly correlated with SFA scores (correlation coefficient approximately 0.75), there are some discrepancies between DEA and SFA efficiency estimates.

9. For either method, it makes little difference whether standardised mortality rates (SMR) or years of life lost (YLL) measures of mortality are used.

10. By replacing the mortality rates for the two disease categories (cancer and circulatory disease) with a single measure of overall mortality it is possible to examine the extent to which some PCTs may be misallocating resources between disease categories (allocating too much to one at the expense of the other). We find very little evidence of any such ‘allocative inefficiency’.

11. The disaggregation of our input measure (total PCT expenditure) into three components (expenditure on cancer, expenditure on circulation problems, and expenditure on all other conditions) has little impact on our findings.

12. We tested many variants of the basic efficiency models described here, including generalisation of the parametric models to a ‘translog’ format. This made little material difference to the conclusions derived from the basic models.

13. We illustrate the potential use of DEA for exploring in more detail the reasons why some PCTs secure low efficiency scores. For example, PCT B secures a DEA efficiency score of 99.5 per cent when compared to a composite PCT comprising a mix of 72.5 per cent of Redcar PCT and 27.5 per cent of Bedfordshire PCT. This composite has the same measured needs as PCT B, a better SMR and 99.5 per cent of the per capita expenditure of PCT B.

14. The final step in the analysis was to examine whether there are characteristics of PCTs that explain the variations in efficiency rankings. Unfortunately, there are few useful measures of PCT structural characteristics. However, we were able to test whether certain indicators of local circumstances were related to efficiency scores. These included:

- the average age of GPs in the PCT
- the proportion of GPs who were female
- the proportion of GPs who secured their degree outside the UK
- the general Index of Multiple Deprivation (IMD) of the PCT
- whether or not the PCT had recently experienced recent merger or reorganisation

- the extent to which the PCT's 2006/07 budget was under or over its target funding allocation
- the scores awarded to the PCT for its use of financial resources (Audit Commission) and its quality of service (Healthcare Commission) for 2006/07
- the scores awarded to the PCT for its commissioning competencies for 2008
- the Quality and Outcomes Framework (QOF) achievement scores recorded by practices within the PCT for eight disease areas in 2005.

15. We find that measured efficiency is associated with several of these factors. Efficiency is strongly negatively correlated with allocations in excess of targets, suggesting that PCTs with relatively generous budgets are less efficient in securing health outcomes than their less generously funded counterparts. This is to be expected. There are likely to be diminishing marginal returns on health expenditure, so overall efficiency is likely to drop as additional funds become available.

16. And, other things being equal, PCTs with higher levels of deprivation tend to be less efficient. We speculate (but cannot confirm) that this may be because such PCTs experience more recruitment difficulties than less deprived PCTs. Or it may be because the needs adjustment used in our analysis does not fully capture the impact of deprivation on health needs. This may be a fruitful area for further research for those developing PCT resource allocation formulae.

17. We also tested the association between efficiency scores and the scores awarded to the PCT by the Audit Commission (for financial management) and the Healthcare Commission (for quality of services). There was a positive correlation between efficiency and a 'good' Audit Commission score and an 'excellent' Healthcare Commission score. We also found some evidence of a positive association between one dimension of commissioning competence – board skills – and efficiency scores.

18. There was also some evidence of a positive association between efficiency levels and the QOF-based COPD achievement score. However, somewhat surprisingly, there was evidence of a negative association between efficiency and the QOF-based diabetes and coronary heart disease achievement scores. This may be because benefits of achievement in these domains do not immediately show through in improved mortality rates.

19. This is a preliminary study of the comparative efficiency of PCTs, measured in the most fundamental way – the extent to which they have purchased good health for their populations given the money they have been allocated. The study has many limitations, not least the limited data available and the fact that it considers only one year's results. WHO experienced similar difficulties with its *World Health Report 2000* (WHO, 2000). However, we believe analysis of this sort is essential if, in the future, PCTs are to be held to account for their performance in commissioning good health for their populations. To the extent that data and methods permit, we have sought to address the admitted limitations and secure results that are consistent across a wide range of specifications. We conclude that there are not major variations in efficiency levels across PCTs, but that there are some anomalies that merit further investigation.

20. From a policy perspective, we believe that it is essential that such efficiency analysis is undertaken as part of the regulatory process, to assure that the NHS is maintaining satisfactory levels of value for money. Although we found generally similar levels of efficiency among PCTs, there were a small number of outlier PCTs that appeared to be achieving significantly lower levels of efficiency than their otherwise identical counterparts, and there is a strong case for further regulatory scrutiny of such PCTs.

21. Finally, it is worth underlining the fact that this analysis examines relative efficiency among PCTs, and therefore says nothing about the overall efficiency with which the NHS commissions health. There may be substantial system-wide initiatives that could improve efficiency across the board, and this study should not lead policy makers to the conclusion that further efficiencies cannot be secured. However, with a few exceptions, seeking to pick out underperforming individual PCTs would not appear to be especially fruitful.

1. Introduction

Expenditure on healthcare accounts for a sizeable proportion of gross domestic product in most developed countries. With an ageing population, increasing pressure on tax revenues and cost increasing health technologies, policy makers are under pressure to ensure that publicly funded healthcare is provided efficiently. International concern with the efficiency of healthcare systems coalesced in the form of the World Health Organization's *World Health Report 2000* (WHO, 2000), which was devoted to the determinants and measurement of health system efficiency at the level of the nation state. Although the report stimulated a wide-ranging international debate and received considerable criticism (Williams, 2001; Anand, Ammar et al, 2002), it did encourage policy makers to focus on the objectives of their health systems, on how achievement might be measured and on whether resources were being deployed efficiently. A subsequent international conference organised by the Organisation for Economic Co-operation and Development confirmed the universal policy concern with performance measurement issues in healthcare (Smith, 2002).

In the UK, Derek Wanless's review of future health spending identified productivity trends as a key determinant of the financial sustainability of the NHS (Wanless, 2002). Peter Gershon was commissioned to undertake a more operational review of public sector efficiency and to identify potential efficiency savings (Gershon, 2004). The Department of Health (DH) was asked to make £2.5bn of efficiency savings in 2005/06, a further £1.9bn of savings in 2006/07, and a further £2.1bn in 2007/08 (DH, 2007a). Thus, by the end of 2007/08, the expected recurring annual efficiency savings were expected to be £6.5bn, which represents about 6 per cent of the Department's budget for 2007/08.

The Office for National Statistics (ONS) has produced a series of reports on productivity trends in the NHS. It estimates that, without any adjustment for quality of care, the quantity of health services provided grew by about 3.9 per cent per annum over the period 2001 to 2005 (ONS, 2008). Inputs grew by 6.5 per cent per annum over the same period, so the ONS estimate of annual productivity change was a decline of 2.5 per cent. However, even after taking account of changes in quality of care (mainly in the form of improvements in post-operative mortality), the estimate of the annual decline in productivity was still 1 per cent.

The concern with productivity has intensified in the wake of the profound economic downturn. A report from the King's Fund and the Institute for Fiscal Studies estimates that, even if the English NHS is protected from real spending cuts, the shortfall from Wanless's most optimistic scenario could range from £21bn to £30bn by 2017 – nearly 30 per cent of the current annual NHS expenditure (Appleby, Crawford and Emmerson, 2009). The authors conclude that major efficiency savings will be required regardless of precise future funding levels. A prime focus for such savings will inevitably be the local NHS commissioning process.

The English NHS is organised geographically, with responsibility for local commissioning devolved to 152 primary care trusts (PCTs). These 152 PCTs have an average population of about 330,000 and controlled about 80 per cent of the NHS's £95bn budget in 2008/09. PCTs are allocated fixed annual budgets by the Department of Health, within which they are expected to commission most aspects of healthcare, including inpatient, outpatient and community care, primary care and pharmaceutical prescriptions. As local organisations, PCTs are considered to be in the best position to understand the needs of their community and are expected to commission the best possible health for their populations, subject to budget constraints. PCTs negotiate contracts for the purchase of healthcare from local providers and they work with local authorities and other agencies that provide health and social care at the local level. For each PCT, data are available for various inputs and outputs, including healthcare expenditure, population demography, the local need for healthcare, and mortality rates. These data can be employed to examine the efficiency with which PCTs convert inputs (in the form of expenditure) into improvements in the health status of their population (in the form of reduced mortality rates), given uncontrollable local environmental conditions (the need for healthcare).

Efficiency concerns the relationship between the outputs produced by an organisation and the inputs it consumes. Economists distinguish two major types of efficiency: technical and allocative. A technically efficient organisation is said to be one that either a) produces the maximum possible volume of outputs given its inputs, or b) produces a specific level of output with the minimum volume of inputs. An allocatively efficient organisation is one that either a) employs the appropriate mix of inputs, given their relative prices, to minimise the cost of producing the chosen level of output, or b) produces the appropriate mix of outputs, given their relative prices, to maximise the value of the chosen level of output. An economically efficient organisation is one that is both technically and allocatively efficient. Overall economic efficiency is measured as the product of an organisation's technical and allocative efficiency. This paper undertakes a preliminary analysis of the technical and allocative efficiency of 152 PCTs.

Two widely employed techniques that generate efficiency measures for a group of similar organisations are econometric analysis and data envelopment analysis (DEA). The econometric analysis group contains two regression methods – corrected ordinary least squares (COLS) and stochastic frontier analysis (SFA) – that can be used to estimate an organisation's efficiency. We examine whether PCTs could reduce their costs and yet still produce the same level of output. Because the cost data will incorporate both technical and allocative inefficiency, the econometric methods can only generate an overall measure of economic efficiency (that is, a measure of technical and allocative efficiency combined). Depending on the model estimated, DEA can provide either a measure of technical efficiency or separate measures of technical and allocative efficiency. By using DEA to estimate a variety of models, we obtain technical and allocative efficiency estimates that can be compared with the economic efficiency estimates from the econometric analysis.

All three methods for measuring efficiency (DEA, COLS and SFA) decompose the measurement of an organisation's efficiency into three steps:

- First, observable phenomena such as inputs and outputs are identified and measured.
- Second, some form of relationship between these phenomena is specified and efficient behaviour is identified (for example, the minimum cost of producing a given set of outputs or the maximum set of outputs that could be produced for any given set of inputs).
- Third, the organisation's actual inputs and outputs are compared with efficient combinations of inputs and outputs, and the difference between these two (or some portion of this difference) is defined as the degree of inefficiency.

However, although DEA and econometric methods (COLS and SFA) share a common approach to efficiency measurement, there are considerable differences between them. For example, COLS and SFA use econometric techniques to estimate the parameters of a specific mathematical form of a cost or production function, while DEA places no conditions on the functional form and uses observed data on inputs and outputs to infer the shape of the frontier. Moreover, while SFA decomposes the unexplained error in the estimated cost or production function into two components – inefficiency (which is always non-negative) and the more conventional two-sided random error term – both COLS and DEA assume that all of the error term is attributable to inefficiency.

The application of these efficiency analysis techniques requires: a) the existence of an adequate number of comparable organisations; b) that the relevant dimensions of performance (such as inputs and outputs) are measurable; and c) that information about the quantities of inputs employed and output produced be readily available to the analyst. In most health systems there are usually several types of organisation, such as hospitals, nursing homes and general practices, each with sufficient numbers to facilitate the application of efficiency analysis techniques; and, because much healthcare activity is publicly funded, there is often no shortage of data. This combination of suitable subjects and data availability has generated considerable academic interest, and both DEA and econometric methods have been widely used in the analysis of efficiency in the healthcare sector.

Hollingsworth (2003) provides a review of the 188 studies published before 2003. He noted that almost two-thirds of these studies were of hospitals and nursing homes in the USA and that almost three-quarters of all studies used DEA alone or DEA plus some analysis of the DEA efficiency ratings. Hollingsworth found that only a small number of studies compared the efficiency ratings generated by different efficiency estimation methods. Five years later, Hollingsworth and Peacock (2008) updated Hollingsworth's original review of the literature. They found 289 studies of efficiency in the health sector, and reported that 48 per cent of these studies used DEA alone, with another 20 per cent using DEA and a second stage regression analysis of the efficiency ratings that sought to explain the causes of efficiency variations. Econometric methods were used by 16 per cent of studies, but only 7 per cent employed both DEA and econometric methods.

This paper contributes to this relatively neglected aspect of the literature. It presents both DEA and statistical efficiency estimates for 152 English PCTs and examines the factors that appear to influence these efficiency ratings.

The following section presents a brief review of efficiency studies related to the UK healthcare sector. In section 3 we outline the major features of the econometric and DEA estimation methods employed in this study, and the differences between the methods are briefly discussed. In section 4 we describe the dataset, outline the models to be estimated and discuss the methods used to estimate them. In section 5 we present and discuss the efficiency scores. In section 6 we present some detailed DEA results for three PCTs; these case studies illustrate the sort of information that DEA can provide for analysts and practitioners. In section 7 we undertake some sensitivity analysis, and in section 8 we attempt to identify those factors that appear to affect PCT efficiency ratings. Section 9 presents some concluding remarks.

2. Efficiency studies of the UK healthcare sector

In their review of 289 studies of efficiency in the health sector, Hollingsworth and Peacock (2008) noted that the emphasis has almost always been on measuring the efficiency of healthcare rather than the efficiency of the production of health for the individual or populations.[1] An exception is the *World Health Report 2000* (WHO, 2000), which sought to measure the efficiency of national health systems in producing population health. Almost 60 per cent of studies examined hospitals and nursing homes in the USA and typically looked at one particular type of hospital (for example, public or private providers). Rather than repeat this review, we focus on the 26 UK studies identified by Hollingsworth and Peacock (2008) and present a brief overview of the main findings.

There have been several studies of hospital efficiency in the UK. Both Hollingsworth and Parkin (1995) and Parkin and Hollingsworth (1997) demonstrated the sensitivity of their DEA results to the variables included in the model for 75 Scottish acute hospitals and found that average efficiency varied considerably across models (from about 80 to 97 per cent). Kerr, Glass et al (1999) used DEA followed by a Tobit regression to estimate the efficiency of 23 acute hospitals in Northern Ireland from 1986/87 and 1991/92. The authors reported that larger units had a mean technical efficiency of 0.94 over the two periods, while smaller units recorded a mean score of 0.91 in 1986/87 and 0.82 in 1991/92.

Jacobs (2001) applied both DEA and SFA methods to data for 232 English hospital trusts for 1995/96. She found that the mean DEA efficiency score varied from 0.64 to 0.936, while the mean SFA score varied from 0.831 to 0.876. She concluded that the differences in efficiency scores across different estimation methods may be due to random noise and data deficiencies. She noted that there were not large efficiency differences between trusts, so improving the poorer performers would generate very modest savings. Street (2003) applied SFA to 226 acute English hospitals using data for 1995/96. He estimated a cost function and obtained mean efficiency levels of 0.694 using COLS, 0.873 using a half-normal SFA and 0.902 using an exponential SFA. Street reported that the two SFA scores were very highly correlated (0.981) and that the COLS and SFA scores were also quite highly correlated (0.831 for the half-normal SFA and 0.905 for the exponential SFA). However, he noted that individual hospital scores were highly sensitive to model specification and cautioned against the use of such scores for hospital-specific performance targets.

There have also been a few UK studies of individual programmes of care. For example, Johnston and Gerard (2001) estimated the efficiency of 64 UK breast screening units in 1996, finding a mean score of 0.821, with large units having a mean of 0.921 and smaller units a mean of 0.845. The authors concluded that the wide variation in efficiency scores across all units, irrespective of their size, may mean that the size of the unit is not significantly related to its efficiency. Buck (2000) used DEA and econometric methods to study 100 community dental services (CDSs) in England in 1997/98. He found that, on average, the CDS was operating at 75 per cent of efficient levels compared to best practice services and that this underperformance could not be explained by factors outside the CDSs' control (such as differences in deprivation and urban–rural differences between localities).

A few studies have examined general healthcare organisations. For example, Salinas-Jiménez and Smith (1996) employed DEA to examine the efficiency of 85 English family health services authorities (FHSAs) for 1991/92 (FHSAs were the administrative unit for primary care at the time of the study). They found that half of the FHSAs were 100 per cent efficient, with the average efficiency level of the inefficient units being 0.926. The authors concluded that there appeared to be little evidence of widespread inefficiency. Giuffrida and Gravelle (2001) used both DEA and COLS/SFA methods to

1 This, of course, is because of the lack of data on the effect of healthcare on the health outcome of the patient.

examine the efficiency of 90 FHSAs for 1993/94 and 1994/95. They found that DEA generated an average efficiency score ranging from 0.91 to 0.99, with COLS generating scores from 0.86 to 0.91 and SFA providing scores from 0.94 to 0.98. They reported very high correlations (over 0.98) between the COLS and SFA FHSA efficiency rankings but noted that the correlations between the rankings from the DEA and COLS/SFA methods were much lower, typically between 0.56 and 0.73.

Several studies have examined efficiency changes through time. For example, Giuffrida (1999) used DEA to calculate indices of productivity change for 90 FHSAs for the years 1990/91 to 1994/95. These so-called Malmquist indices are decomposed into indices of technical efficiency change, scale efficiency change and technological change. Giuffrida found a small improvement in overall efficiency. Average technical efficiency was 0.9961 in 1990/91 and 0.9995 in 1994/95. This implies that the potential technical efficiency gain is very small and that almost all of this gain was realised by the end of the study period. Giuffrida also reported a small positive change in scale efficiency but no significant technological gain. Maniadakis and Thanassoulis (2000) used data from 75 Scottish acute hospitals from 1991/92 to 1995/96 to estimate Malmquist indices of efficiency gain. The authors reported a small pure technical efficiency gain (rising from 0.86 in 1991/92 to 0.89 in 1995/96), with a larger allocative efficiency gain (which rose from 0.77 to 0.84). With scale efficiency unchanged at 0.90, overall efficiency rose from 0.60 to 0.69, which implies a large degree of inefficiency across the sector.

It is clear from the literature that: a) neither DEA nor econometric methods are considered to be superior – they are seeking to measure subtly different concepts of efficiency; and b) the use of different input and output specifications can generate markedly different results. Researchers employ either or both methods, and in the latter compare the results that they generate. Analysts also estimate a variety of models to examine the impact of different specifications. Accordingly, we utilise both DEA and COLS/SFA to estimate the efficiency of English PCTs and we estimate a variety of models (with different inputs and outputs) to examine the sensitivity of the results to both model specification and estimation method.

Furthermore, as described further in section 4, we have available three disease-specific indicators (the mortality rate from cancer, mortality from circulation problems and mortality from all other causes) to use as output measures and a number of input measures, including total PCT expenditure on healthcare per head of population. This allows us to examine, albeit imperfectly, the total 'health system' efficiency, measured in terms of the health conferred on PCT populations, thereby distinguishing our study from the predominantly 'healthcare' analyses described above.

3. Approaches to efficiency measurement

In this section we give an outline of the major approaches to efficiency measurement when the analyst is only able to observe each organisation for a single year. Other techniques are available when data are available for more than one point in time but, as our dataset is solely cross-sectional, these other techniques are not discussed here.[2] This section is relatively technical and draws heavily on material presented in Jacobs, Smith and Street (2006). It can be safely skipped by those interested only in the substantive results of the study.

3.1 Background: production and cost functions

As econometric-based efficiency estimates are usually a by-product from the estimation of a production or cost function, we need to say a little about such functions. Many studies of healthcare organisations adopt a neo-classical approach. Here, the production function summarises a technical relationship between the maximum output attainable for different combinations of all possible factors of production. For example, where hospital output is measured as the total number of patients treated (*Y*) and there are two factors of production, labour (*L*) and capital (*K*), the production function would be written as follows:

Equation 1

$$Y = f(L, K)$$

where *f*(.) describes the functional relationship between output and different mixes of labour and capital.

One of the most widely used production functions is the Cobb-Douglas, which takes a logarithmic form and can be written as follows:

Equation 2

$$Y = \alpha L^{\beta_1} K^{\beta_2}$$

and estimated as follows:

Equation 3

$$\ln Y_i = \alpha + \beta_1 \ln L_i + \beta_2 \ln K_i + \varepsilon_i$$

where $\beta1$ and $\beta2$ are parameters describing the contributions to output made by labour and capital respectively. The logarithmic form allows these parameters to be interpreted as elasticities: a 1 per cent increase in the amount of labour employed is predicted to lead to a percentage increase in output equal to the value of $\beta1$.

Another commonly estimated production function is the transcendental logarithmic (translog) function (Christensen, Jorgenson and Lau, 1973). The attraction of the translog is its flexibility, as it can approximate virtually any functional form (Intriligator, 1978). The translog function is estimated by adding

2 Although expenditure data for these PCTs is available for 2004/05, 2005/06 and 2006/07, outcome data are only available for one time period (that is, pooled for the three-year period 2004/06).

the squares and cross-products of the explanatory variables to the Cobb-Douglas function. Thus the translog version of equation 3 would be estimated as:

Equation 4

$$\ln Y_i = \alpha + \beta_1 \ln L_i + \beta_2 \ln K_i + \frac{1}{2}\beta_3(\ln L_i)^2 + \frac{1}{2}\beta_4(\ln K_i)^2 + \beta_5 \ln L_i \ln K_i + \varepsilon_i$$

If the parameters $\beta 3$, $\beta 4$ and $\beta 5$ are not significantly different from zero, the translog function reduces to a Cobb-Douglas function.

However, in many industries, estimation of a production function poses serious practical difficulties (Intriligator, 1978). In particular, where organisations produce multiple outputs, it is a challenge to derive a composite measure of output without loss of information. Faced with multiple outputs, most researchers find it more convenient to work with a cost function because it allows a single dependent variable, cost (C), to be estimated. Information about different outputs can be included as a vector of explanatory variables. If cost-minimising behaviour can be assumed, the cost function is usually the dual of the production function, making the two approaches equivalent.[3]

The cost function equivalent to the production function of equation (1) can be written as:

Equation 5

$$C = f(Y, w, r)$$

where w and r represent input prices for labour (wages) and capital (rent) respectively.

The cost function equivalent to the Cobb-Douglas production function in equation 2 is as follows:

Equation 6

$$C(Y, w, r) = \alpha'(Y w^{\beta_1} r^{\beta_2})^{1/(\beta_1+\beta_2)}$$

The elasticities of $\beta 1$ and $\beta 2$ can be estimated from a linear model of the following form:

Equation 7

$$\ln C_i = \alpha + \frac{1}{\beta_1+\beta_2}\ln Y_i + \frac{\beta_1}{\beta_1+\beta_2}\ln w_i + \frac{\beta_2}{\beta_1+\beta_2}\ln r_i + \varepsilon_i$$

where w and r are the unit price of each factor of production.

The translog cost function can be estimated as follows:

3 The assumption of cost-minimising behaviour (that is, that there is no allocative inefficiency) is unlikely to be met in practice, so efficiency estimates will incorporate both allocative and technical inefficiency.

Equation 8

$$\begin{aligned}\ln C_i &= \alpha + \beta_0 \ln Y_i + \beta_1 \ln w_i + \beta_2 \ln r_i \\ &+ \frac{1}{2}\beta_3 (\ln Y_i)^2 + \frac{1}{2}\beta_4 (\ln w_i)^2 + \frac{1}{2}\beta_5 (\ln r_i)^2 \\ &+ \beta_6 \ln Y_i \ln w_i + \beta_7 \ln Y_i \ln r_i + \beta_8 \ln w_i \ln r_i + \varepsilon_i\end{aligned}$$

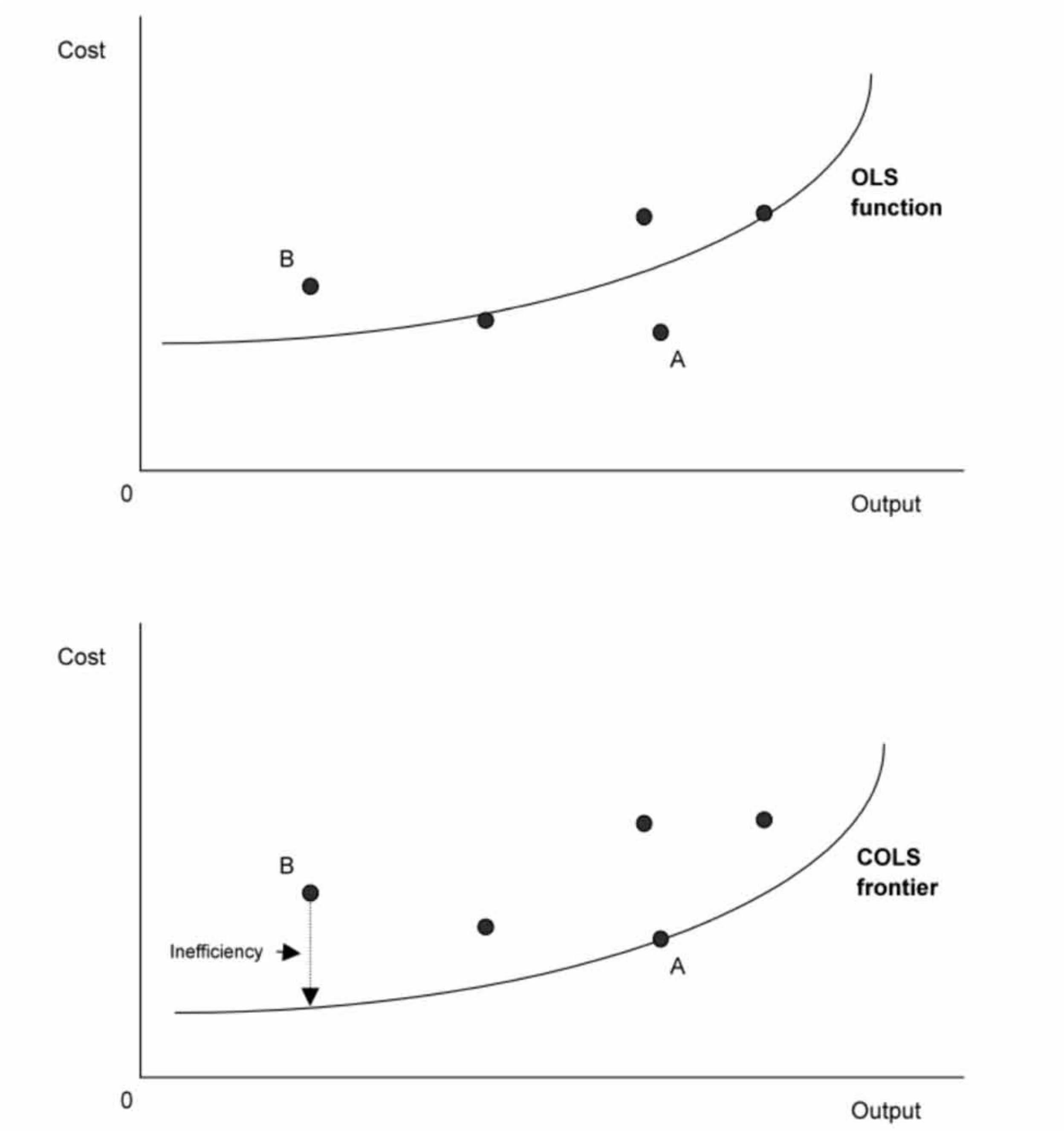
This will correspond to the translog production function of equation 4 if factor markets are competitive and the cost function displays constant returns to scale, with total costs increasing proportionally when all prices increase proportionally, given the level of output (Christensen and Greene, 1976).

3.2 Rudimentary econometric analysis (corrected OLS)

Estimation of a cost function such as that in equation 8 provides information about the average relationship between costs and a set of explanatory variables for the organisations within the sample, but it does not directly reveal ‘best practice’. However, Farrell (1957) suggested that the residuals from such a regression could be used to describe the extent to which an organisation deviates from best practice. For example, the residuals from a cost function show the extent to which actual costs differ from predicted costs. Thus an organisation with a zero residual could be interpreted as one with an average level of efficiency, and the organisation with the most negative residual could be considered to be the most efficient. This interpretation implies that a cost frontier can be estimated by shifting the OLS regression line downwards until it just passes through the organisation with the most negative residual.

Figure 1 illustrates this procedure for a single explanatory variable (say, output) regressed on costs. The upper figure shows the fitted OLS function through the set of observations. Under the COLS approach, the organisation with the most negative residual is defined as being fully efficient (its costs are lower than those for any other organisation holding constant the other variables in the model). The COLS efficiency frontier is located by shifting the OLS regression line downwards so that it passes through this fully efficient observation. This is illustrated in the lower half of figure 1, where observation A is efficient. For an organisation lying above the COLS frontier, it is predicted that it would be able to reduce costs to the level predicted by the best practice frontier without having to reduce output.

Figure 1: Illustration of an OLS regression and COLS frontier



Efficiency ratings for each organisation can be calculated by estimating the parameters of the cost function by OLS and then shifting down the intercept until all residuals are non-negative and one residual is equal to zero. This is achieved by adding the most negative residual ($\min(\varepsilon_i)$) to the intercept and subtracting $\min(\varepsilon_i)$ from all of the residuals so that the OLS regression line shifts downwards. Instead of passing through the centre of a cloud of observed data, the regression line now passes through the single observation displaying minimum cost. The COLS residuals (εc_i) are calculated as $\varepsilon c_i=\varepsilon_i-\min(\varepsilon_i)$. If a logarithmic cost function has been estimated such as that in equation 7 or 8, the efficiency values are calculated as $\exp(\varepsilon c_i)$ and fall between 1 and infinity. For reporting purposes, it is usual to invert these values so that the efficiency scores are presented as $1/\exp(\varepsilon c_i)$ and lie between 0 and 1. These scores represent the percentage distance from the frontier. In this approach, the entire residual is attributed to inefficiency and the most efficient unit is defined as 100 per cent efficient.

3.3 Stochastic frontier analysis (SFA)

Although the COLS approach implies that the residual reflects only inefficiency, another viewpoint would attribute the entire residual to random influences or measurement error. Alternatively, the residual might comprise both of these elements – inefficiency and random (stochastic) error – and stochastic frontier analysis (SFA) has been developed to provide separate estimates of these two components.

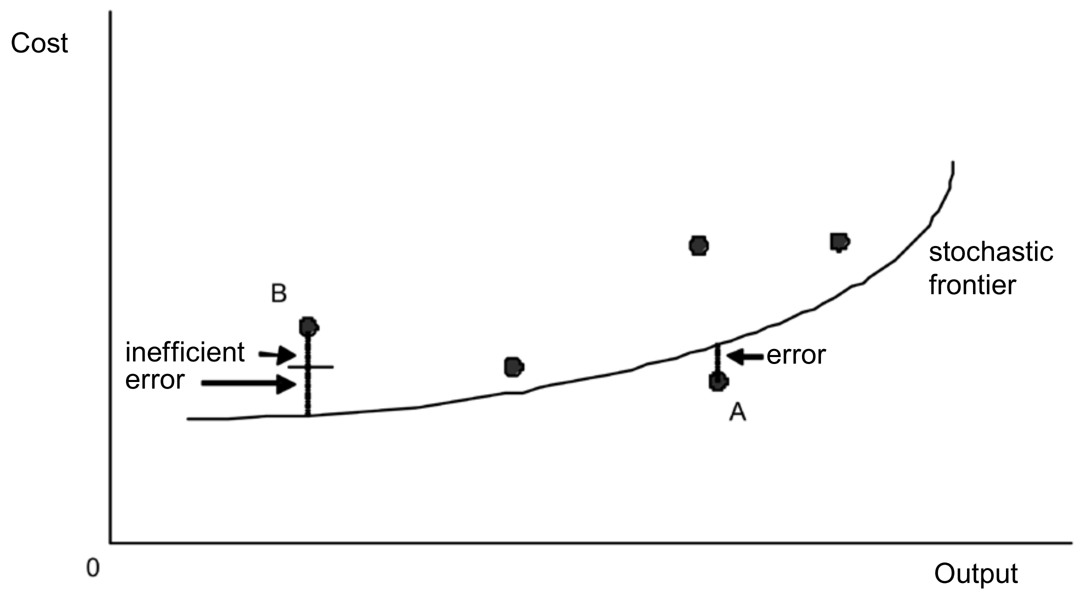
The key assumption underlying SFA is that the inefficiency component and the random component of the residual are distributed differently. In particular, the random component is assumed to be distributed normally, as is consistent with the classical OLS model. If ε_i (for example, from equation 7 or 8) is normally distributed, all residual variance is interpreted as arising from random noise and measurement error. If ε_i is skewed, this is taken as evidence that there is inefficiency in the sample. Subject to ε_i being skewed, stochastic frontier analysis decomposes the error term into two parts with zero covariance as follows:

Equation 9

$$\varepsilon_i = v_i + u_i, \qquad \operatorname{cov}(v_i, u_i) = 0 \qquad u_i \geq 0$$

The dual specification is defended on the grounds that each component represents an economically distinct disturbance. The v_i can be interpreted as stochastic (random) events not under the control of the organisation, whereas u_i is a non-negative term that captures the cost of inefficiency, with u_i defining how far the organisation operates above the cost frontier. Diagrammatically, this might result in a cost function similar to that depicted in figure 2. The frontier does not (necessarily) pass through the observation that has lowest cost because the frontier is estimated after recognising that the difference between observed cost and the level of cost predicted by the explanatory variables is not due solely to inefficiency. Some of the difference may be due to measurement error and omitted variables. In figure 2, observation A lies below the estimated frontier. The distance of this point from the stochastic cost frontier is attributable to random error, v_i. For observations lying above the frontier, the distance comprises both measurement error and inefficiency, as illustrated for observation B.

Figure 2: Illustration of a stochastic frontier



To estimate the stochastic frontier it is necessary to specify the distributional characteristics of the two components of the residual term. The random error term (v_i) is typically assumed to be normally distributed with a zero mean and constant variance. No economic criteria are available to guide the choice of the distribution of the inefficiency term (u_i), but standard computer software allows several options including the half-normal, truncated (at zero) normal, and exponential distributions. Jacobs, Smith and Street (2006) show that the choice of distribution may affect the inefficiency estimates (for example, the exponential distribution will impose a highly skewed relationship so that most observations are clustered close to the frontier with a long tail of observations further away).

As was the case for COLS, SFA efficiency values, calculated as $\exp(u_i)$ for a cost function with logged variables, fall between 1 and infinity. For reporting purposes, it is usual to invert these values so that the efficiency scores are presented as $1/\exp(u_i)$ and lie between 0 and 1. Again, these scores represent the percentage distance from the frontier but, unlike the COLS approach, only part of the residual term is attributed to inefficiency and the most efficient unit is not necessarily 100 per cent efficient.

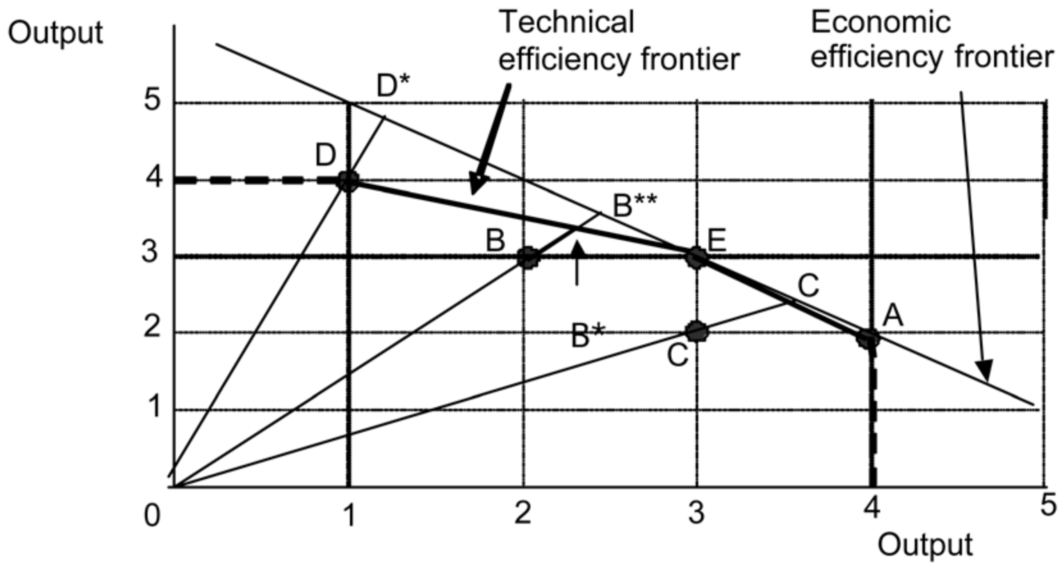
3.4 Data envelopment analysis (DEA)

While the COLS/SFA approach is informed by economic theory, DEA is purely data-driven. The location and the shape of the efficiency frontier are determined by the data. The frontier is based on the notion that an organisation that employs less input than another organisation to produce the same amount of output can be considered more efficient. Those observations with the highest ratios of output to input are considered efficient, and the efficiency frontier is constructed by joining up these observations in the input/output space. The frontier thus comprises a series of linear segments connecting one efficient observation to another. The construction of the frontier is based on ‘best observed practice’ and is therefore only an approximation to the true unobserved efficiency frontier. Inefficient organisations are ‘enveloped’ by the efficiency frontier in DEA and the inefficiency of the organisations within the frontier boundary is calculated relative to this surface.

To illustrate these ideas, suppose that there are five organisations (labelled A, B, C, D and E) using one input to produce two outputs as depicted in figure 3 (with the outputs measured in terms of their volume not value). Figure 3 shows the outputs produced from each unit of input for each organisation. The technical efficiency frontier is the piecewise series of bold (solid line) linear segments with the (dashed line) horizontal and vertical extensions. Organisations D, E and A all exhibit 100 per cent technical efficiency as they lie on (and indeed help to form) the frontier. However, B lies within the frontier and has a technical efficiency rating indicated by the ratio of OB/OB*. Its efficient comparators are D and E, and a weighted average of the outputs of D and E (at B*) would produce more output than B but with the same level of input. Similarly, C is inefficient because a weighted average of E and A (at C*) would produce more output than C but with the same level of input.

Although D, E and A are all 100 per cent technically efficient – no other organisation or combination of organisations produces more of both outputs per unit of input than they do – not all are likely to be allocatively efficient (that is, produce the greatest value of output given the relative prices of the outputs). If we assume that both outputs have the same price then we can draw a straight line economic efficiency frontier through the co-ordinates (1,5), (2,4), (3,3), (4,2) and (5,1). Organisations E and A are on this frontier and are thus both technically and allocatively efficient. Although D is technically efficient, it is producing the wrong mix of outputs given their relative prices and it exhibits some allocative inefficiency indicated by the ratio OD/OD**. Although C is technically inefficient, it is producing the correct mix of outputs and so exhibits no allocative inefficiency. B is both technically inefficient (indicated by OB/OB*) and allocatively inefficent (indicated by OB*/OB**), with an overall economic efficiency of OB/OB**.

Figure 3: Illustration of a DEA frontier (output orientation)



Efficiency in DEA is defined as the ratio of the weighted sum of outputs of a decision-making unit (DMU) divided by a weighted sum of its inputs. It is computed by solving for each DMU the following mathematical programme:

Equation 10

$$\max \left(\frac{\sum_{s=1}^{S} u_s \times y_{s0}}{\sum_{m=1}^{M} v_m \times x_{m0}} \right)$$

subject to:

$$\frac{\sum_{s=1}^{S} u_s \times y_{si}}{\sum_{m=1}^{M} v_m \times x_{mi}} \le 1 \qquad i = 1,\ldots, I$$

where:

y_{s0} = quantity of output s for DMU_0

u_s = weight attached to output s, $u_s > 0$, $s = 1,\ldots, S$

x_{m0} = quantity of input m for DMU_0

v_m = weight attached to input m, $v_m > 0$, $m = 1,\ldots, M$

This mathematical programme seeks out for DMU_0 the set of output weights u_s and input weights v_m that maximise the efficiency of DMU_0, subject to the important constraint that, when these weights are applied to all other DMUs, no DMU can have an efficiency greater than 1. The weights can take any non-negative value, and in general a different set of weights is computed for each DMU. Thus the weights

u_s and v_m are a central feature of DEA. They are chosen to cast each DMU in the 'best possible light', in the sense that no other set of weights will yield a higher level of efficiency. In creating the efficient frontier, DEA yields specific input or output targets for each DMU, depending on whether an input or output orientation has been specified. For example, under input orientation, the input targets indicate the specific amounts by which a particular DMU should be able to reduce its consumption of particular inputs without reducing output.

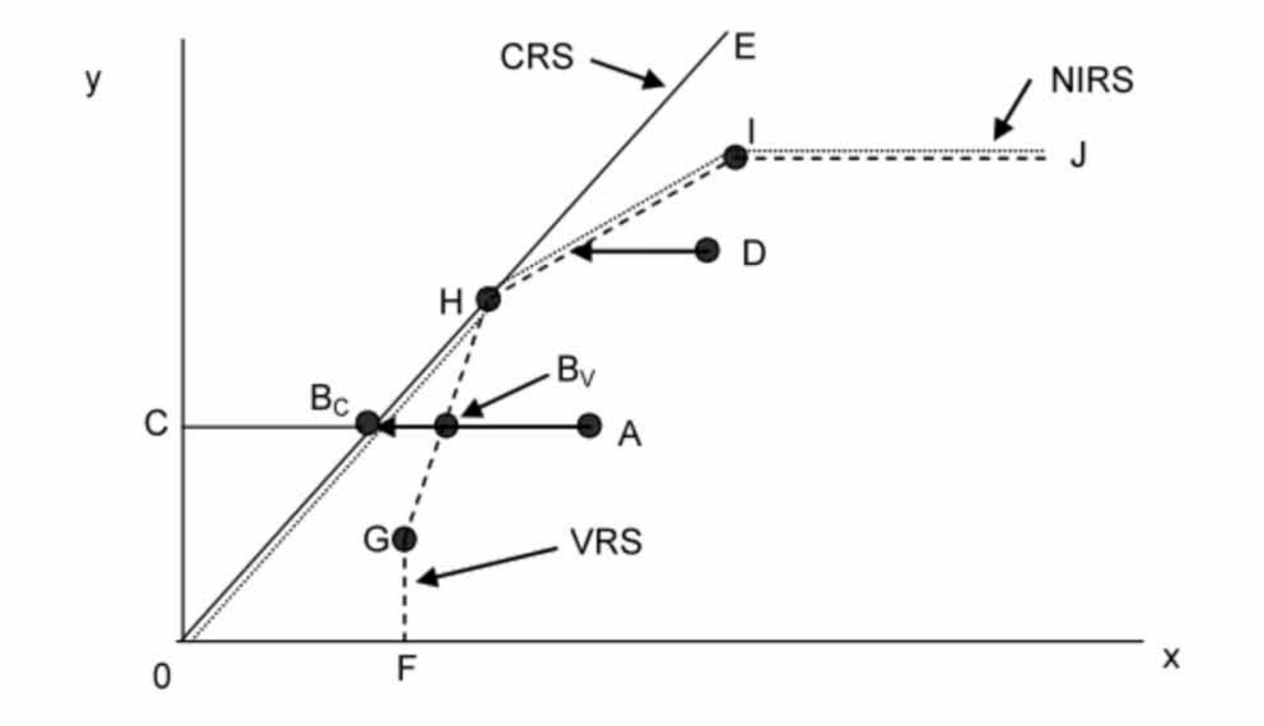
The above description assumes constant returns to scale (CRS). However, an organisation may find itself operating at an inefficient scale for reasons outside the control of its managers and, in such a situation, efficiency estimates based on the assumption of constant returns would be too low. To correct this, the model needs to be estimated with variable returns to scale (VRS). This is illustrated in the following example. Assume organisation A produces a single output (*y*) from a single input (*x*) as depicted in figure 4. The line from the origin OE depicts the CRS frontier, whereas the linear segmented frontier FGHIJ is the VRS frontier. Assuming an input orientation, implying a reduction of input (*x*) in the horizontal plane, the *technical efficiency* ($TE_{IN,CRS}$) of A with respect to the constant returns to scale technology is then expressed as follows:

Equation 11

$$TE_{IN,CRS} = \frac{CB_C}{CA}$$

where the *IN* subscript denotes the input orientation and the *CRS* subscript denotes the constant returns to scale technology.

Figure 4: Constant and variable returns to scale (input orientation)



In contrast, the *technical efficiency* ($TE_{IN,CRS}$) of A with respect to the variable returns to scale technology is expressed as follows:

Equation 12

$$TE_{IN,VRS} = \frac{CB_V}{CA}$$

where the *VRS* subscript denotes the variable returns to scale technology. These efficiency measures are bounded by 0 and 1.

The VRS frontier shown in figure 4 envelops the data points, such as A and D, more tightly than the CRS approach, where the frontier would extend from the origin. Thus, by introducing an additional constraint, VRS produces technical efficiency scores which are greater than or equal to those obtained using CRS. Under VRS, an inefficient organisation will usually be compared only with organisations of a roughly similar size. Thus, the projected point for A on the DEA frontier will be a convex combination of other organisations such as G and H. This convexity restriction implies that the efficient frontier is formed only by interpolation between organisations and precludes extrapolation of performance at one scale to a different scale. In contrast, the CRS case permits extrapolation, with the result that organisations may be compared with others operating at substantially different scales.

Being a non-parametric technique, DEA has the advantage of requiring no assumptions about the functional form of the production or cost frontier. While this reduces the need for a theoretical exposition of model specification, it does not avoid the problem of how to assess the quality of a DEA model or how well it reflects reality and these concerns lead to several issues. First, DEA is deterministic, which means there is no way to take account of statistical error, random shocks or noise. Given that the method is based on outlier observations, measurement error is a potentially serious source of bias. The approach presupposes that all variables are measured accurately and that any shortfall between a DMU's input/output ratio and the maximum predicted by the frontier is attributable solely to inefficiency. Measurement error can have an impact on both the DMU's own efficiency rating and the rating for other DMUs for which it is a peer. It has been suggested that the DEA inefficiency score is likely to contain measurement error and that it may be best to consider it as an equivalent to the efficiency score from the COLS model.

Second, DEA results are sensitive to model specification, particularly in small samples. DEA generates efficiency scores for each individual organisation by comparing it to peers that produce a comparable mix of outputs. If any output is unique to an organisation, it will have no peers with which to make a comparison, irrespective of the fact that it may produce other outputs in common. An absence of peers results in the automatic assignation of full efficiency for the DMU under consideration and this may reflect an unusual input/output mix rather than full efficiency.

Third, the analyst using DEA has to be clear about what variables should be classified and included as inputs to or outputs of the production process. There is no agreed method to determine whether or not a variable should be included in the model. Generally, the criteria of *exclusivity* and *exhaustiveness* should hold for the choice of inputs and outputs in a DEA model (Thanassoulis, 2001). In other words, the inputs alone must influence the outputs (*exclusivity*) and only those outputs used in the model (*exhaustiveness*). The inputs and outputs therefore need to be chosen such that the inputs capture all resources and the outputs capture all activities or outcomes deemed as relevant for the particular efficiency analysis, subject to the rule of exclusivity and exhaustiveness. The exclusion of an important input or output can result in severely biased results and an underestimate of efficiency because it may fail to recognise input constraints faced by some DMUs. Conversely, the addition of extraneous input or output in DEA will tend to lead to overestimates of efficiency scores because an unnecessary constraint has been added into the linear programme. The bias, however, tends to be much more modest for including an extraneous variable than omitting a relevant variable (Smith, 1997).

Fourth, there is the issue of how to adjust for environmental variables. Such variables describe factors that could influence the efficiency of a DMU but are not traditional inputs to the production process and are assumed to be outside the control of the manager. Often environmental variables are included as one of the inputs in the production model, so that DMUs are only compared with other DMUs operating in identical or more adverse environments. Those operating in the most adverse environments will automatically be deemed efficient. However, Jacobs, Smith and Street (2006) note that the literature provides several different recommendations on how to handle such variables. Consequently, there is no generally accepted single method for incorporating environmental variables into DEA models or for testing whether an environmental variable has a significant influence on the production process and any resulting efficiency estimates.

Finally, sensitivity analysis may help to refine the model specification. DEA offers no diagnostic statistics with which to judge whether a model is misspecified. Analysts should therefore test a variety of model specifications under sensitivity analysis to ascertain the robustness of results.

3.5 COLS, SFA and DEA compared

Having outlined the main features of each estimation method, we now briefly examine why they might generate different estimates of organisational efficiency.[4] Jacobs, Smith and Street (2006) suggest that there are two main reasons for these differences and these relate to:

- differences in how the techniques establish and shape the efficiency frontier, and
- differences in how the techniques determine how far individual observations lie from the frontier.

COLS/SFA is based on the theory of the firm but requires assumptions to be made about the functional form of the cost/production function. DEA is not based on any theoretical underpinning but requires no assumptions about the underlying structure of production. In DEA, the frontier is defined solely by the data: the outermost observations, given the scale of operation, are defined as efficient. As such, the frontier is positioned and shaped by the data, not by theoretical considerations. The drawback of this, however, is that the location of the DEA frontier is sensitive to observations that may have unusual types, levels or combinations of inputs and outputs. These will have a scarcity of adjacent reference observations or 'peers', perhaps resulting in sections of the 'frontier' being unreliably estimated and inappropriately positioned.

Another drawback with DEA concerns how the technique interprets any distance from the frontier. There are two key differences between DEA and SFA. First, DEA assumes correct model specification and that all data are observed without error. SFA (but not COLS) allows for the possibility of modelling and measurement error. Consequently, if the two methods yield an identical frontier, SFA efficiency estimates are likely to be higher than those produced by DEA. Second, DEA uses a selective amount of data to estimate individual efficiency scores. DEA generates efficiency scores for each organisation by comparing it only to peers that produce a comparable mix of outputs. This has two implications. First, if any output is unique to an organisation, it will have no peers with which to make a comparison, irrespective of the fact that it may produce other outputs in common. An absence of peers results in the automatic assignation of full efficiency to the organisation under consideration. Second, when assigning an inefficiency score to an observation lying away from the frontier, only its peers are considered, with information pertaining to the remainder of the sample discarded. In contrast, SFA appeals to the full sample information when estimating relative efficiency. In addition to making greater use of the available

4 We assume that the DEA and COLS/SFA models are such that they permit the estimation of the same type of efficiency (for example, either they both estimate only technical efficiency or they both estimate technical and allocative efficiency).

data, this facet of the estimation procedure will make individual efficiency estimates more robust in the presence of outlier observations and to the presence of atypical input/output combinations.

One of the key strengths of DEA over COLS/SFA is that it can readily model multiple output production processes. COLS/SFA cannot readily handle these, but can do so if a cost function rather than a production function is estimated. Both methods may be susceptible to the influence of outliers and small sample sizes. DEA is more vulnerable to outliers because of its inherent process of ‘placing each organisation in the best possible light’. As such, organisations with unusual production processes can easily be promoted to the efficiency frontier. Because COLS/SFA estimates are derived from full sample information, the technique is less prone to outlier influence. Of course, it may be that ‘outliers’ are the very organisations that are most inefficient, so excluding them on the basis of statistical criteria may undermine the exercise altogether. Small sample sizes do not prevent the application of DEA, but, as with all parametric estimation processes, COLS/SFA estimates are likely to be more imprecise the smaller the sample size.

4. Data, models and methods

In this section we outline the dataset, the efficiency models to be estimated, and the methods employed to estimate these models.

4.1 Data

The unit of analysis is the PCT. Data for the efficiency analysis have been extracted from two principal sources: the Department of Health's national programme budgeting project, and outcome indicators calculated by the National Centre for Health Outcomes Development (NCHOD). Since April 2003, each PCT has been required to allocate all of its expenditure to one of 23 broad programmes of care, or programme budgeting categories (PBCs), such as cancer (PBC 2), circulation problems (PBC 10), respiratory problems (PBC 11) and gastro-intestinal problems (PBC 13). These programmes are defined with reference to the International Classification of Diseases (ICD) version 10 codes at the four-digit level, and most programme budget categories reflect broad ICD 10 chapter headings. This programme budgeting dataset embraces most items of publicly funded expenditure, including inpatient, outpatient and community care, and pharmaceutical prescriptions. For this study we focus on expenditure in three programme budgeting categories: cancer, circulation problems, and all other categories combined. The descriptive statistics in table 1 show that total PCT expenditure averaged £436m in 2006/07 with, on average, £27m spent on cancer and £40m spent on circulation problems in each PCT.

Table 1: Descriptive statistics for the data

Variable	Mean	Std Dev	Min	Max
Total expenditure (2006/07, £)	436,533,956	219,418,712	132,715,000	1,395,694,976
Cancer expenditure (2006/07, £)	26,855,270	15,783,210	7,937,000	103,532,000
Circulation expenditure (2006/07, £)	40,222,243	22,269,419	12,109,000	145,154,000
Other expenditure (2006/07, £)	369,456,445	183,047,768	111,880,000	1,147,009,024
Population 2006/07	332,080	186,401	90,121	1,258,847
Population 2006/07, MFF adjusted	332,080	186,456	84,482	1,290,843
Population 2006/07, MFF and need adjusted	332,080	166,336	100,650	1,093,330
Need for health care index 2006/07	1.029	0.138	0.716	1.400
Total spend per head (MFF adjusted, £	1,351	178	992	1,926
Cancer spend per head (MFF adjusted, £)	81	18	42	151
Circulation spend per head (MFF adjusted, £)	124	26	67	200
Other spend per head (MFF adjusted, £)	1,145	150	835	1,619
Direct SMR cancer, aged <75 years	120	14	76	165
SYLL rate cancer, aged <75 years	158	18	103	218
Direct SMR circulation problems, aged <75yrs	90	19	55	142
SYLL rate circulation problems, aged <75yrs	108	25	65	177
Direct SMR all causes of death, aged <75yrs	326	56	211	495
SYLL rate all causes of death, aged <75yrs	483	83	318	742
Direct SMR amenable causes of death, aged <75yrs	118	23	69	186
SYLL rate amenable causes of death, aged <75yrs	153	30	88	249

Notes

1. The descriptive statistics for all variables are unweighted.
2. The expenditure and population data are for the financial year 2006/07.
3. The SMR and SYLL rate variables are based on pooled mortality data for the three-year period 2004/06.
4. For the definition of those causes of death deemed amenable to healthcare see Martin, Rice and Smith (2008).

Sources: Calculated by the authors from data in the programme budget section of the Department of Health's website at www.dh.gov.uk and from data in the compendium indicators (mortality) section of the NCHOD website at www.nchod.nhs.uk (December 2007 version).

The programme budget dataset also includes an estimate of the population for each PCT. This averages 332,000, ranging from 90,000 in the smallest PCT to over 1,250,000 in the largest PCT. Some PCTs face considerably higher input prices in the local health economy than other PCTs (for example, input prices are up to 40 per cent higher in London and the south-east of England than elsewhere) and those PCTs facing higher input prices receive a proportionately higher budget. By multiplying the raw population figure for each PCT by an index reflecting local input prices and then rescaling these estimates so that they sum to the total population of England, we obtain the Market Forces Factor (MFF) adjusted population.[5] If we divide total expenditure by the MFF adjusted population, we obtain an estimate of PCT expenditure per head of population, which is adjusted for local variations in input prices.

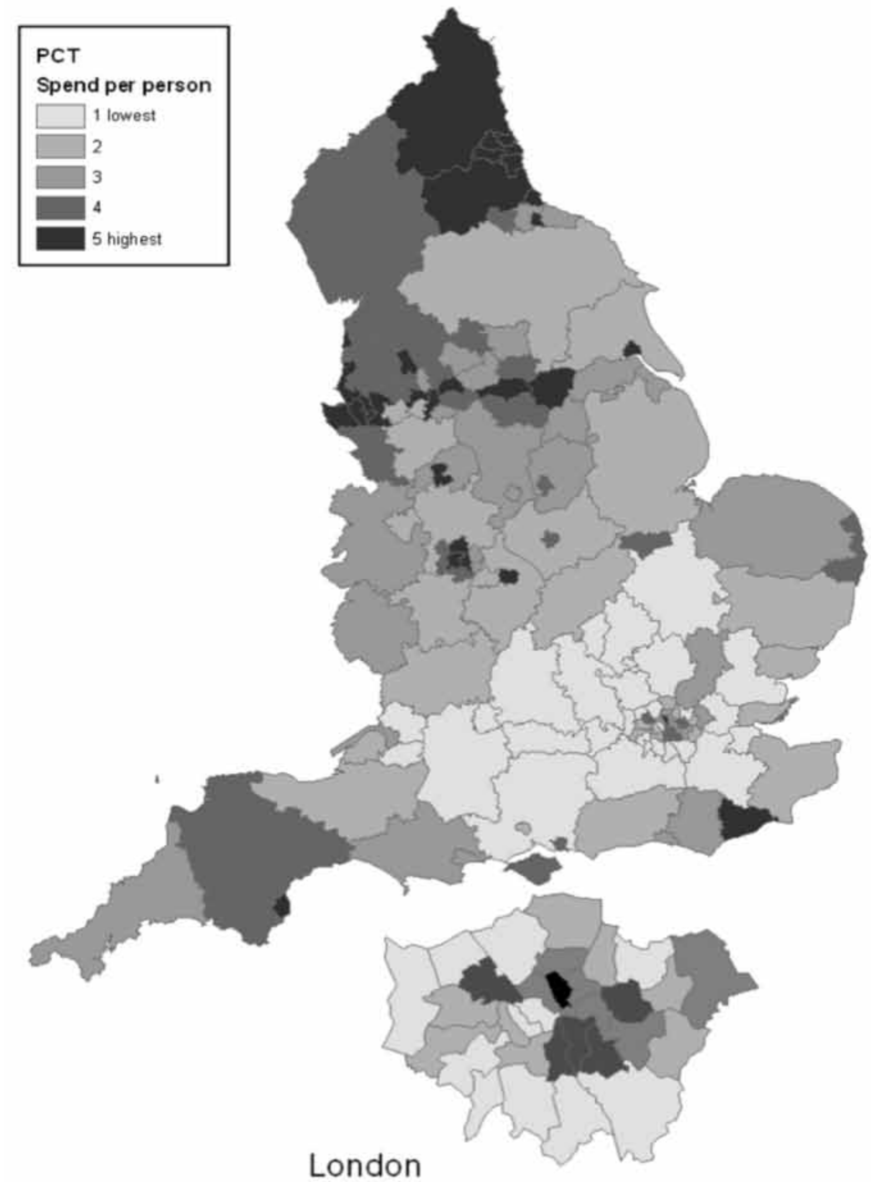
To illustrate the geographic variation in expenditure levels, figure 5 shows expenditure per person (adjusted for local input prices) by PCT for 2006/07, with each PCT allocated to one of five quintiles (quintile 1 contains the 30 PCTs with the smallest spend per head and quintile 5 contains the 30 PCTs with the largest spend per head).[6] This shows that, per capita, expenditure is typically greatest around London and the traditional industrial heartlands in the north-east, the north-west, West Yorkshire, and the West Midlands. Table 1 shows that expenditure per head (adjusted for local input prices) averages £1,351 across all PCTs but varies between £992 and £1,926 per person. Similarly, the average spend per head on cancer (adjusted for local input prices) is £81 but this varies between £42 and £151, and the average spend per head on circulation problems averages £124 but this varies between £67 and £200.

The need for healthcare is unlikely to be constant across PCTs. Areas with relatively large proportions of elderly residents or PCTs operating in relatively deprived locations can be expected to have a relatively high need for healthcare. The Department of Health recognises this by adjusting per capita budgetary allocations to PCTs according to a complex 'needs' formula, derived from an econometric analysis of the link between healthcare expenditure and socio-economic factors at a small area level within England (DH, 2005). The programme budgeting dataset also reports a 'unified weighted' population for each PCT which reflects its raw population adjusted for local input prices, the demographic profile and the local need for healthcare generated by socio-economic conditions. We divide the unified weighted population (which incorporates an MFF adjustment) by the MFF adjusted population to obtain an index of the (demographic and socio-economic) need for healthcare. This 'need for healthcare' index averages about 1 across all PCTs but varies substantially, with some PCTs having a needs index 30 per cent below the national average and others having a need for healthcare 40 per cent above the national average.

5 We have used the Market Forces Factor (MFF) indices that feed into the payment by results tariffs for 2007/08 to adjust expenditure for local input prices (see DH, 2007b).

6 We would like to thank Stephen Clark of the City Development Directorate, Leeds City Council, and Peter Halls of the University of York Computing Service for their assistance with the construction of the PCT maps. Both maps are Crown Copyright 2007. All rights reserved. Ordnance Survey Licence number 100018355.

Figure 5: Expenditure per person by PCT, 2006/07



The mortality data employed in this study were released by NCHOD in December 2007 and relate to deaths over the three-year period 2004/06. Figure 6 uses an age and sex standardised mortality rate (SMR) to allocate PCTs to one of five mortality quintiles (quintile 1 contains the 30 PCTs with the smallest SMR and quintile 5 contains the 30 PCTs with the largest SMR).[7] Mortality rates seem to be greatest in those areas with the largest expenditure.

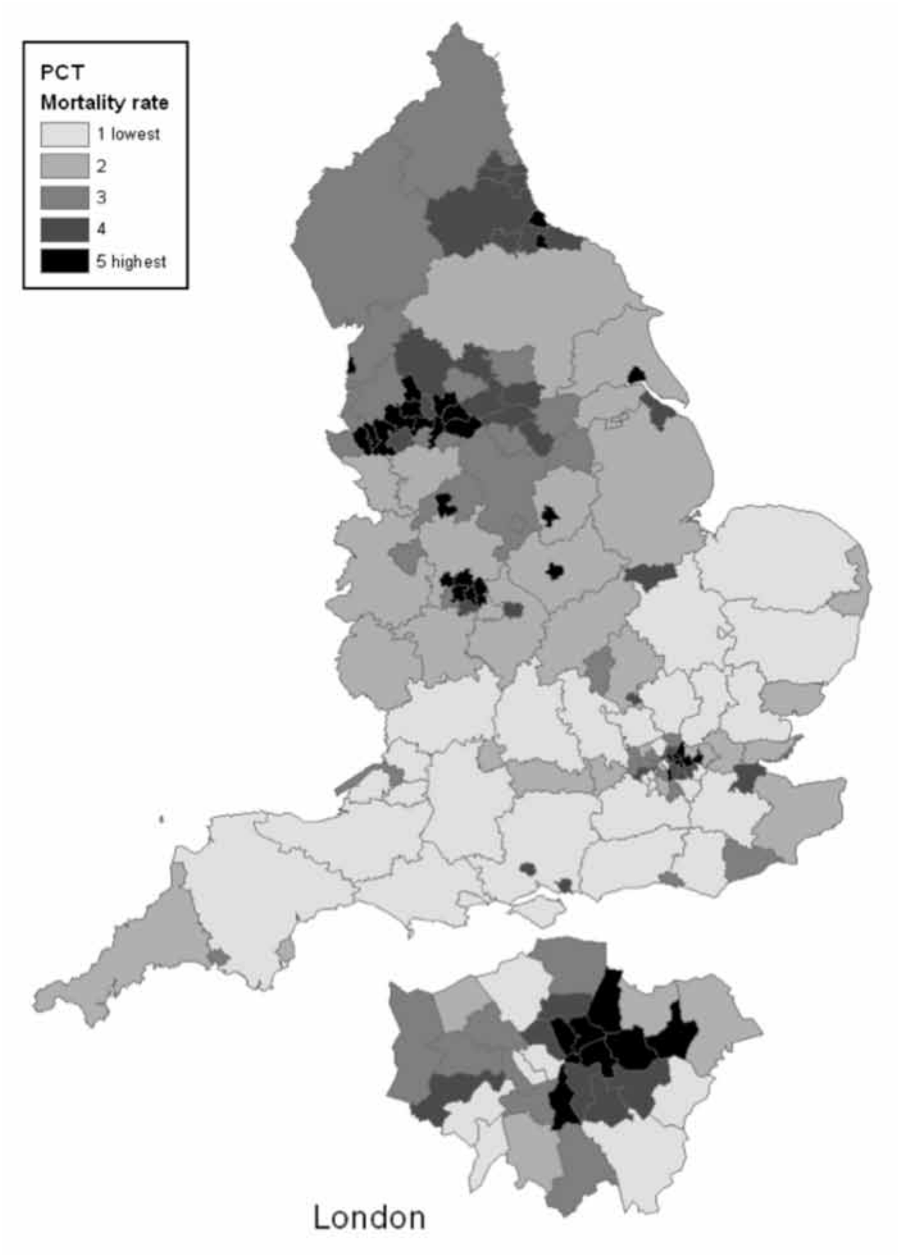
Table 1 shows that the direct (age) SMR for those aged under 75 years and dying from cancer averages 120 deaths per 100,000 population across all PCTs, but this varies between 76 deaths and 165 deaths per 100,000 population. Similarly, the direct SMR for those aged under 75 years and dying from circulation problems averages 90 deaths per 100,000 population annually, and this varies between 55 and 142 deaths per 100,000 population. The direct SMR for those aged under 75 years and dying from

7 The SMR is for people aged under 75 years and relates to all causes of death considered amenable to healthcare over the period 2004/06.

all causes averages 326 deaths per 100,000 population, while the same SMR from all causes deemed amenable to healthcare is 118 deaths per 100,000 population.

In addition to the SMRs, we also employ a measure of the avoidable years of life lost (YLL). This is calculated by summing over ages 1 to 74 years the number of deaths at each age multiplied by the number of years of life remaining up to age 75 years. The crude YLL rate is simply the number of years of life lost divided by the resident population aged under 75 years. Like conventional mortality rates, YLL can be age standardised to eliminate the effects of differences in population age structures between areas, and this (age) standardised YLL (SYLL) rate is the second health outcome variable employed in this study (Lakhani, Olearnik and Eayres, 2006: 379). As table 1 shows, on average, 158 years of life were lost annually to cancer per 10,000 population over the three-year period 2004/06. For circulation problems, 108 years of life were lost each year per 10,000 population. For all causes of death, 483 years of life were lost annually, and 153 years of life per 10,000 population were lost annually for deaths from causes considered amenable to healthcare.

Figure 6: Mortality rate by PCT, 2004/06



The descriptive statistics presented in table 1 are all for variables in absolute form. However, DEA requires that outputs should be measured in such a way that more output is considered to be better than less output. Consequently, when estimating the DEA models with mortality as an output, we use the inverse of the mortality rates. For the econometric analysis we use the natural logarithm of the value for each variable.

4.2 Models

The purpose of this study is not so much to find the 'best' fitting model but to examine the sensitivity of the efficiency scores to the estimation method employed and the model fitted, and to try to identify those factors that affect efficiency levels. Table 2 summarises the three basic models and their variants to be estimated, as well as the estimation techniques employed.

Model 1

The first model estimated (model 1) uses expenditure per head of population (adjusted for local input prices) as the input variable in DEA and as the dependent variable in COLS/SFA. Three mortality indicators – for cancer, circulation problems and all other causes – are included as DEA outputs and as regressors in the COLS/SFA approach. In addition, we also include the need for healthcare as an uncontrollable DEA output and as an additional regressor in the COLS/SFA regressions. Two variants of this model are estimated: model 1a uses SMRs as the mortality indicators, while model 1b uses SYLL rates as outputs (for DEA) or regressors (for COLS/SFA). The DEA results for this model provide technical efficiency scores, while the econometric efficiency estimates reflect both technical and allocative efficiency combined (that is, economic efficiency).

Model 2

Model 2 is similar to model 1 but replaces the three separate mortality indicators with a single measure. Four variants of this model are estimated, each with a different mortality indicator:

- model 2a uses the all-deaths SMR
- model 2b uses the all-deaths SYLL rate
- model 2c uses the SMR for deaths from causes considered amenable to healthcare
- model 2d uses the SYLL rate for deaths from causes considered amenable to healthcare.

As was the case for model 1, the econometric methods generate efficiency scores that reflect technical and allocative efficiency combined (that is, economic efficiency). However, by amalgamating the three mortality rates into a single indicator (by giving each death the same weight irrespective of its cause), the DEA scores reflect total economic efficiency levels and are directly comparable with the COLS/SFA scores. If we divide the economic efficiency ratings for models 2a and 2b by the technical efficiency ratings for models 1a and 1b, we obtain allocative efficiency scores for models 1a and 1b.

A standard neo-classical cost function such as that in equation 7 includes factor prices as regressors so that the cost implications of choosing a particular production process would be captured by the model parameters, not by the residual. However, factor prices have been omitted from our regressor set partly because, within the NHS, factor prices are set nationally through central bargaining processes and partly because local input price variations have been incorporated through the appropriate adjustment of the dependent variable (that is, by adjusting the denominator of the dependent variable – the size of the population – for the so-called Market Forces Factor (MFF)).

Another rationale for not including capital and labour prices as explanatory variables would be that the amount and mix of inputs is determined by hospital managers, so any sub-optimality arising from the employment of these resources should be considered as indicative of inefficiency. Such a 'behavioural' cost function formulation would only include those cost-influencing factors over which hospitals have no control, so that their influence on costs can be eliminated when estimating efficiency levels. Cost differences remaining over and above those 'explained' by the behavioural econometric model are deemed to reflect differences in organisational effort and choice about what technical production process to employ. By omitting factor prices and all other variables endogenous to managerial influence, their impact will be captured by the residual term and incorporated into the efficiency estimates (Giuffrida, Gravelle and Sutton, 2000).

Model 3

Model 3 is similar to model 2, with a mortality indicator and the need for healthcare as outputs. However, the single input in model 2 – expenditure per head (adjusted for local input prices) – is now disaggregated into three parts: expenditure per head on cancer, expenditure per person on circulation problems, and expenditure per head on all other categories of healthcare. With multiple inputs and multiple outputs we cannot readily estimate this model using an econometric approach, but we can still obtain technical efficiency ratings from DEA and compare these ratings with those obtained for models 1 and 2.

4.3 Methods

As some of the variables in the dataset are ratios (for example, healthcare expenditure per person), the DEA models are estimated assuming a variable returns to scale (VRS) technology with an input orientation. The use of VRS is essential with variables expressed as ratios (Hollingsworth and Smith, 2003), and we focus on an input orientation because managers are likely to have more control over expenditure (which we treat as an input) than over either local mortality rates or the need for healthcare (which we treat as an uncontrollable environmental factor on the output side). The resulting efficiency estimates show by how much each PCT could reduce its expenditure and still produce the same output given the same environmental conditions. However, for comparative purposes, we also report results for an output maximisation orientation for the first two models. In this scenario, the efficiency estimates show by how much each PCT could reduce its mortality rates and/or operate with an increased local need for healthcare without increasing the current level of inputs.

The COLS efficiency estimates are relatively straightforward to derive: we estimate an OLS regression and then shift the regression line downwards so that it passes through the observation with the most negative residual. However, to obtain the SFA efficiency estimates, we need to specify the distribution of the inefficiency term. We report efficiency estimates for three distributions: the half-normal, the truncated (at 0) normal, and the exponential distribution. Descriptive statistics for each set of efficiency ratings are presented together with a coefficient reflecting the degree of correlation between any two sets of ratings.

Table 2: Model specifications

Model	DEA			COLS/SFA		
	Outputs estimated	Inputs estimated	Efficiency	Regressand	Regressors	Efficiency
1a	Cancer SMR Circulation SMR Other SMR Need for healthcare	Cost per head	Technical	Cost per head	Cancer SMR Circulation SMR Other SMR Need for healthcare	Economic
1b	Cancer SYLL rate Circulation SYLL rate Other SYLL rate Need for healthcare	Cost per head	Technical	Cost per head	Cancer SYLL rate Circulation SYLL rate Other SYLL rate Need for healthcare	Economic
2a	All-causes SMR Need for healthcare	Cost per head	Economic	Cost per head	All-causes SMR Need for healthcare	Economic
2b	All-causes SYLL rate Need for healthcare	Cost per head	Economic	Cost per head	All-causes SYLL rate Need for healthcare	Economic
2c	Amenable causes SMR Need for healthcare	Cost per head	Economic	Cost per head	Amenable causes SMR Need for healthcare	Economic
2d	Amenable causes SYLL Need for healthcare	Cost per head	Economic	Cost per head	Amenable causes SYLL rate Need for healthcare	Economic
3a	All-causes SMR Need for healthcare	Cancer cost per head Circulation cost per head Other cost per head	Technical	n/a		
3b	All-causes SYLL rate Need for healthcare	Cancer cost per head Circulation cost per head Other cost per head	Technical	n/a		
3c	Amenable cause SMR Need for healthcare	Cancer cost per head Circulation cost per head Other cost per head	Technical	n/a		
3d	Amenable cause SYLL Need for healthcare	Cancer cost per head Circulation cost per head	Technical	n/a		

Notes

1. All mortality indicators are for people aged under 75 years of age.
2. COLS/SFA cannot be readily used to estimate model 3 because this model has multiple inputs and multiple outputs.
3. Economic efficiency=technical efficiency*allocative efficiency

5. Results

In this section we report the efficiency scores for the various models and estimation methods outlined in table 2. The models that generate these efficiency scores are of less interest; we are primarily interested in the efficiency scores. Details of the statistical models can be found in the appendix. The distribution parameters for both the half-normal and exponential SFA models (lambda and theta respectively) are significant for all models estimated, suggesting that these models are an improvement on OLS estimation (Street, 2003). The distribution parameter for the truncated normal model (mu) is not significant but the results for this distribution are reported for the sake of completeness.

5.1 Results for model 1

Table 3 provides descriptive statistics for the efficiency scores from both variants of model 1. For model 1a, the average efficiency score for the three SFA models exceeds that for both COLS and DEA. This is to be expected as SFA partitions the error term into inefficiency and random error, whereas both COLS and DEA assume that there is no random error and that any excess cost is attributable to inefficiency.[8] The efficiency scores for the three SFA models appear to be very similar, with almost identical means, variances, and minimum and maximum values. The efficiency scores for the two DEA models (input minimisation and output maximisation) are also very similar.

The replacement of the mortality rates (model 1a) with YLL rates (model 1b) has very little impact on the average efficiency score and its variance. The average SFA score declines very slightly but the average COLS/DEA efficiency score increases by a similarly small amount (and the number of efficient PCTs increases by 2 from 16 to 18). The PCTs with the highest output to expenditure ratio for each output will be automatically deemed technically efficient by DEA because there are no other PCTs with which to compare them. In addition to these four PCTs, DEA identifies another dozen or so PCTs as being 100 per cent efficient in models 1a and 1b.[9]

8 This assumes that any allocative inefficiency is small relative to the technical efficiency (remember that the DEA efficiency estimates for model 1 are only for technical efficiency while the COLS/SFA estimates are for technical and allocative efficiency combined). As we will see, the allocative efficiency estimates implied by models 1a and 2a and by models 1b and 2b suggest that the average allocative efficiency rating is over 0.99.

9 To ensure that the outcome measures increase as performance improves, we have used the inverse of the relevant mortality rates as our outcome measures. However, the way in which the outcome measure is adjusted can affect the DEA results. Thus, in addition to using the inverse of the mortality rates, we also re-estimated the two DEA models 1a and 1b using mortality rates that had been subtracted from 1,000. However, the technical efficiencies generated by this approach were very similar to those using the inverse of the mortality rates: the correlation coefficient for the two sets of (input minimisation) efficiencies was 0.991 for model 1a and 0.994 for model 1b.

Table 3: Model 1 efficiency scores

Model 1a	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores (I)	DEA technical efficiency scores (O)
	Half-normal	Truncated	Exponential			
Mean	0.961	0.971	0.971	0.880	0.921	0.945
Std dev	0.020	0.019	0.019	0.039	0.051	0.036
Min	0.884	0.881	0.881	0.767	0.776	0.830
Max	0.990	0.992	0.992	1.000	1.000	1.000
Efficient PCTs					16	16

Note: Model 1a includes a single input (expenditure per head) and three SMRs and need as outputs. DEA (I)=DEA input minimisation. DEA (O)=DEA maximisation.

Model 1b	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores (I)	DEA technical efficiency scores (O)
	Half-normal	Truncated	Exponential			
Mean	0.957	0.967	0.969	0.885	0.922	0.950
Std dev	0.023	0.021	0.021	0.040	0.052	0.034
Min	0.873	0.874	0.875	0.772	0.776	0.826
Max	0.990	0.991	0.991	1.000	1.000	1.000
Efficient PCTs					18	18

Note: Model 1b includes a single input (expenditure per head) and three SYLL rates and need as outputs.

The 152 PCTs in this study vary considerably in size (from 93,000 patients to 1,265,000 patients) and we checked for any impact of scale on the technical efficiency scores by splitting the PCTs into five groups according to size. As table 4 shows, there is no obvious relationship between the average efficiency score and PCT size quintile.

Table 4: Average efficiency score and size of PCT

PCT size	Average efficiency score (all for model 1a)		
	SFA HN	COLS	DEA (input min)
Qunitile 1 (=smallest PCTs)	0.963	0.889	0.933
Quintile 2	0.957	0.872	0.910
Quintile 3	0.959	0.877	0.899
Quintile 4	0.959	0.876	0.921
Quintile 5 (=largest PCTs)	0.965	0.885	0.942

Note: Each quintile consists of 30 PCTs except the fifth quintile, which contains 32 PCTs.

Figure 7 shows the distribution of the SFA half-normal and DEA efficiency scores for model 1a. These confirm the descriptive statistics reported above, with the DEA scores exhibiting a much wider dispersion than the SFA scores and the DEA scores being, on average, lower than the SFA scores.

Figure 7: SFA half-normal and DEA (input minimisation) efficiency scores for model 1a

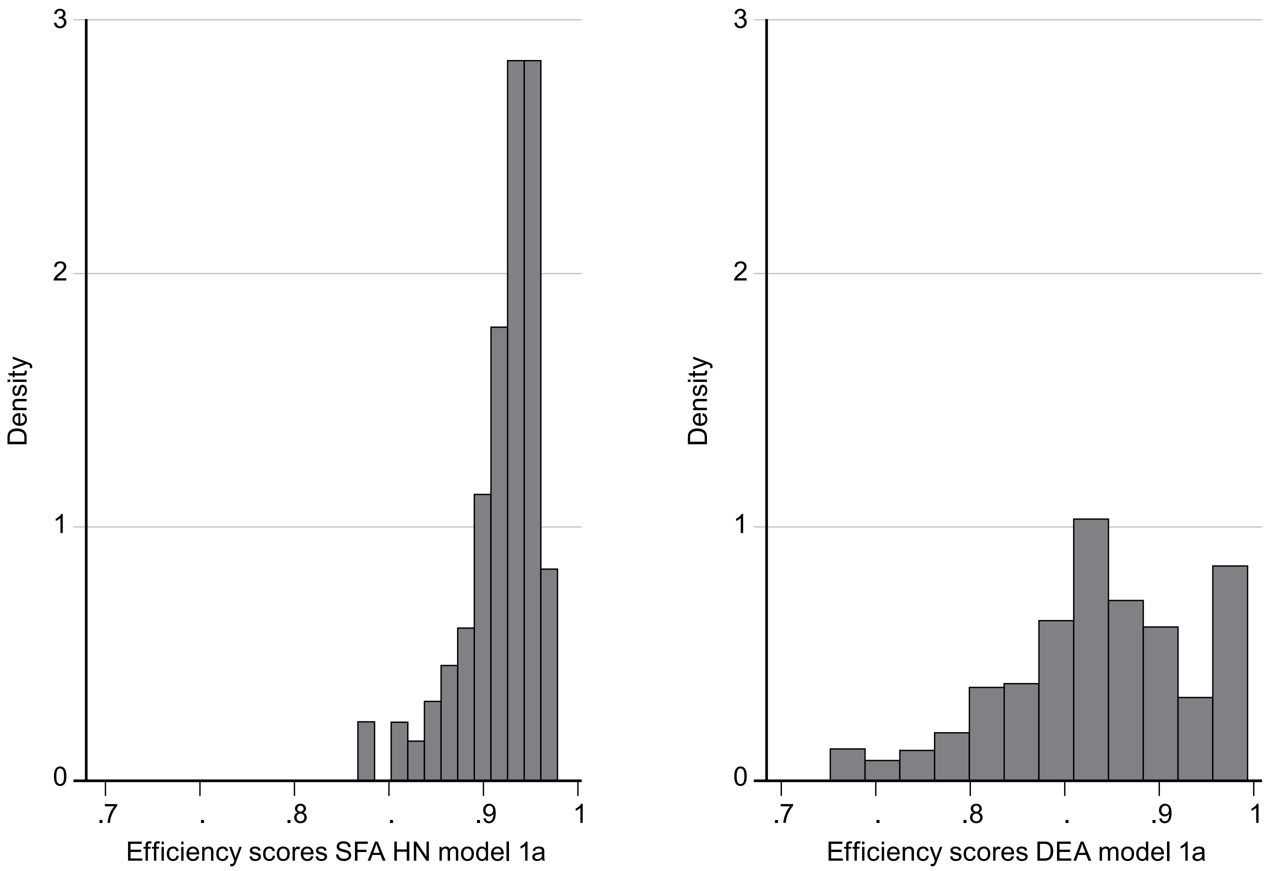


Table 5 shows the correlations between the various efficiency scores for model 1a. As is to be expected, the three SFA models are extremely highly correlated with each other (with correlation coefficients of more than 0.98) and each SFA model is highly correlated with the COLS model (with a coefficient of between 0.88 and 0.95). The input minimisation DEA efficiency scores are slightly less highly correlated with the scores from the econometric approaches but even here the correlations are still considerable at about 0.75. The output maximisation DEA scores are less well correlated with the scores from the econometric approach, with a correlation coefficient of about 0.60.

Table 5: Correlation matrix of DEA and COLS/SFA efficiency scores for model 1a

	Half-normal	Truncated	Exponential	COLS	DEA input min	DEA output max
Half-normal	1					
Truncated	0.982	1				
Exponential	0.981	0.999	1			
COLS	0.947	0.884	0.881	1		
DEA input min	0.774	0.753	0.751	0.749	1	
DEA output max	0.604	0.568	0.565	0.596	0.873	1

Table 6 shows the correlations between the various efficiency scores for model 1b. Again, the three SFA models are extremely highly correlated with each other (with a correlation coefficient of more than 0.98) and each SFA model is highly correlated with the COLS model (with a coefficient of about 0.90). Both the input minimisation and output maximisation DEA efficiency scores are less highly correlated with the scores from the econometric approaches for model 1b: the correlation coefficient declines to about 0.72 and 0.50 respectively.

Table 6: Correlation matrix of DEA and COLS/SFA efficiency scores for model 1b

	Half-normal	Truncated	Exponential	COLS	DEA input min	DEA output max
Half-normal	1					
Truncated	0.988	1				
Exponential	0.981	0.999	1			
COLS	0.951	0.905	0.890	1		
DEA input min	0.733	0.728	0.724	0.704	1	
DEA output max	0.510	0.500	0.494	0.490	0.854	1

Table 7 reports the individual PCT scores and rankings for model 1a for 20 PCTs. These confirm the impression generated by the descriptive statistics and correlations. The efficiency scores for all three variants of the SFA model are very similar, and although the COLS and SFA scores differ, the PCT rankings for COLS and SFA are also very similar. Although there is some difference between the DEA and COLS/SFA rankings, PCTs at the top and bottom of the rankings for DEA (for example, PCTs G and N respectively) are also at the top and bottom of the rankings for COLS/SFA. The PCTs that change rank the most seem to be those in the middle of the rankings; this finding is also reported by Street (2003) in his efficiency analysis of English hospitals.

Table 7: Individual PCT scores and rankings for model 1a

PCT	Efficiency scores (model 1a) DEA (input minimisation)	DEA (output maximisation)	COLS	SFA/HN	SFA/TN	SFA/E	PCT	Efficiency rankings (model 1a) DEA (input minimisation)	DEA (output maximisation)	COLS	SFA/HN	SFA/TN	SFA/E
A	0.917	0.949	0.888	0.969	0.979	0.979	A	68	83	86	88	87	87
B	0.997	0.997	0.981	0.988	0.991	0.991	B	135	135	150	150	150	150
C	0.946	0.954	0.867	0.96	0.973	0.973	C	104	96	54	54	55	55
D	0.951	0.973	0.891	0.971	0.98	0.98	D	108	118	94	94	94	94
E	0.929	0.947	0.914	0.978	0.984	0.985	E	86	81	127	126	126	126
F	0.966	0.973	0.883	0.968	0.978	0.979	F	123	120	77	82	83	83
G	1	1	0.961	0.986	0.99	0.99	G	152	152	148	148	148	148
H	1	1	0.93	0.981	0.986	0.987	H	152	152	142	140	141	141
I	1	1	0.941	0.983	0.988	0.988	I	152	152	145	144	144	144
J	0.963	0.972	0.893	0.972	0.981	0.981	J	119	117	98	101	101	101
K	0.89	0.935	0.883	0.965	0.976	0.976	K	37	51	75	68	68	68
L	0.983	0.984	0.966	0.987	0.99	0.99	L	129	126	149	149	149	149
M	0.91	0.948	0.907	0.975	0.982	0.982	M	56	84	118	113	113	113
N	0.792	0.951	0.774	0.889	0.887	0.887	N	2	88	3	2	2	2
O	0.94	0.959	0.929	0.981	0.986	0.986	O	98	101	141	141	140	140
P	0.967	0.973	0.92	0.979	0.985	0.986	P	126	119	134	136	135	135
Q	0.875	0.911	0.873	0.961	0.973	0.973	Q	27	24	60	58	57	57
R	0.883	0.947	0.801	0.915	0.926	0.927	R	31	79	7	7	7	7
S	0.912	0.923	0.92	0.978	0.984	0.985	S	60	38	133	128	127	127
T	0.873	0.89	0.86	0.953	0.967	0.968	T	25	9	40	38	38	38

Note: The efficiency rankings are from 1 (=least efficient) to 152 (=most efficient).

Table 8 shows the correlations between the efficiency scores for selected versions of models 1a and 1b. This shows that the choice of mortality or YLL rates has little impact on the efficiency scores, with extremely high correlations (over 0.95) for a given estimation technique. As was noted above, the correlations are slightly lower across different estimation techniques.

Table 8: Correlation matrix for selected efficiency scores from models 1a and 1b

	Half-normal 1a	Half-normal 1b	COLS 1a	COLS 1b	DEA (I) 1a	DEA (I) 1b	DEA (O) 1a	DEA (O) 1b
Half-normal 1a	1							
Half-normal 1b	0.993	1						
COLS 1a	0.947	0.939	1					
COLS 1b	0.945	0.951	0.990	1				
DEA (I) 1a	0.774	0.780	0.749	0.755	1			
DEA (I) 1b	0.733	0.733	0.704	0.704	0.978	1		
DEA (O) 1a	0.604	0.615	0.596	0.600	0.873	0.861	1	
DEA (O) 1b	0.507	0.510	0.495	0.490	0.824	0.854	0.957	1

Note: DEA (I)=DEA input minimisation. DEA (O)=DEA maximisation.

From this point on we focus on the DEA results using an input orientation, which indicates the potential savings that a locality could make if it were to perform at the same level of efficiency as its analogous peers with the same (or more adverse) needs and the same (or better) mortality rates. That is, the efficiency estimates show by how much each PCT could reduce its expenditure and still produce the same mortality rates given the same local need for healthcare.

5.2 Results for model 2

Table 9 provides descriptive statistics for the efficiency scores for the four variants of model 2. These models have expenditure per person as the single input/regressand, and a mortality indicator (the mortality or YLL rate for either all causes of death or causes amenable to healthcare) and the need for healthcare as the outputs/regressors. Although the DEA efficiency scores for model 1 only reflect technical efficiency (because weights for the different outputs have not been specified), the scores for model 2 reflect total economic (technical and allocative) efficiency (because the separate mortality indicators have been combined into a single indicator, with each death carrying the same weight irrespective of its cause). Thus the DEA efficiency scores for model 2 are directly comparable with the COLS/SFA scores for model 2. Moreover, if we divide the DEA economic efficiency ratings for models 2a and 2b by the technical efficiency ratings for models 1a and 1b, we obtain allocative efficiency ratings for the latter two models.

Descriptive statistics for the scores for all four variants of model 2 are presented in table 9 and are very similar to those for model 1 in table 3. The average efficiency score for the three SFA models exceeds that for both COLS and DEA. The efficiency scores for the three SFA models appear to be very similar, with almost identical means, variances, and minimum and maximum values. The average DEA efficiency rating appears to have declined slightly, although this is to be expected because the ratings in table 9 reflect both technical and allocative inefficiency whereas the DEA ratings in table 3 only reflect technical inefficiency. For the same reason, the number of 100 per cent efficient PCTs also declines (from 16 in table 3 to 11 in table 9). If we divide the DEA economic efficiency ratings for models 2a and 2b by the technical efficiency ratings for models 1a and 1b, we obtain allocative efficiency ratings for the latter two models; these average 0.993 and 0.991 respectively.[10] This result implies that there is very little allocative inefficiency and that most inefficiency is technical inefficiency.

Table 9: Model 2 efficiency scores

Model 2a	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores
	Half-normal	Truncated	Exponential		
Mean	0.959	0.970	0.970	0.883	0.915
Std dev	0.021	0.020	0.020	0.040	0.050
Min	0.874	0.870	0.870	0.764	0.776
Max	0.989	0.991	0.991	1.000	1.000
Efficient PCTs					11

Note: Model 2a includes a single input (expenditure per head) with an all-causes SMR and need as outputs.

Model 2b	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores
	Half-normal	Truncated	Exponential		
Mean	0.959	0.970	0.970	0.885	0.915
Std dev	0.020	0.020	0.020	0.040	0.051
Min	0.879	0.874	0.874	0.769	0.776
Max	0.989	0.991	0.991	1.000	1.000
Efficient PCTs					11

Note: Model 2b includes a single input (expenditure per head) with an all-causes SYLL rate and need as outputs.

Model 2c	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores
	Half-normal	Truncated	Exponential		
Mean	0.956	0.968	0.968	0.882	0.914
Std dev	0.024	0.022	0.022	0.041	0.049
Min	0.862	0.861	0.861	0.759	0.776
Max	0.990	0.991	0.991	1.000	1.000
Efficient PCTs					11

Note: Model 2c includes a single input (expenditure per head) with an amenable to healthcare SMR and need as outputs.

10 The corresponding minimum allocative efficiency ratings are 0.925 (for model 1a) and 0.926 (for model 1b). Curiously, the maximum allocative efficiency rating for both models is marginally above 1 (1.00088 for model 1a and 1.00086 for model 1b), with model 1a (1b) recording 7 (1) PCTs with an efficiency rating greater than unity.

Model 2d	SFA economic efficiency scores			COLS economic efficiency scores	DEA technical efficiency scores
	Half-normal	Truncated	Exponential		
Mean	0.955	0.967	0.967	0.882	0.915
Std dev	0.253	0.023	0.023	0.041	0.049
Min	0.857	0.857	0.857	0.759	0.776
Max	0.990	0.992	0.992	1.000	1.000
Efficient PCTs					11

Note: Model 2d includes a single input (expenditure per head) with an amenable to healthcare SYLL rate and need as outputs.

Figure 8 shows the distribution of the SFA half-normal and DEA efficiency scores for model 2a. These are similar to those shown in figure 7 for model 1a and demonstrate that models 1a and 2a generate similar results. As was noted for model 1a, the two graphs in figure 8 show that the DEA scores exhibit greater dispersion than the SFA scores, and that the DEA scores are on average lower than the SFA scores.

Figure 8: SFA half-normal and DEA efficiency scores for model 2a

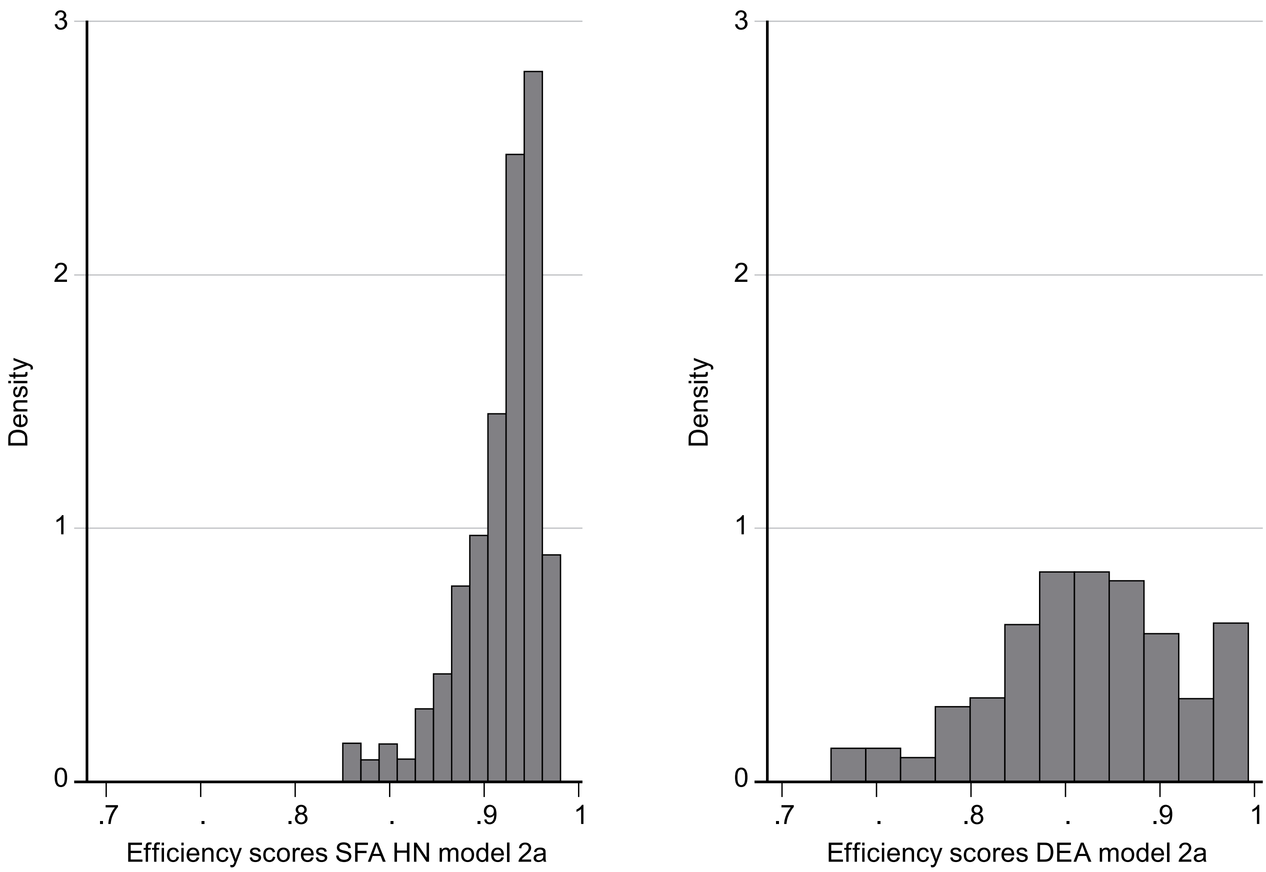


Table 10 shows the correlations between the efficiency scores for the five estimation methods applied to model 2a. As was noted for model 1, the three SFA models are extremely highly correlated with each other (indeed, the correlation between the efficiency ratings for the exponential and truncated normal models is 1) and each SFA model is highly correlated with the COLS model (with a coefficient of about 0.88 to 0.95).[11] The DEA efficiency scores are less highly correlated with the scores from the econometric approaches but even here the correlation coefficient is still considerable at about 0.80. The correlations between the efficiency scores using the same five estimation methods for models 2b, 2c and 2d are very similar to those for model 2a (and are not presented here).

Table 10: Correlation matrix of DEA and COLS/SFA efficiency scores for model 2a

	SFA half-normal	SFA truncated	SFA exponential	COLS	DEA
SFA half-normal	1				
SFA truncated	0.979	1			
SFA exponential	0.979	1	1		
COLS	0.947	0.880	0.880	1	
DEA	0.817	0.783	0.783	0.794	1

To illustrate, table 11 reports the individual PCT scores and rankings for model 2a for 20 PCTs that illustrate a spectrum of circumstances. As was the case for model 1a, these individual scores and rankings confirm the impression generated by the descriptive statistics and correlations. The efficiency scores for all three variants of the SFA model are very similar, and although the COLS and SFA scores differ, the PCT rankings for COLS and SFA are very similar. A comparison of the scores and rankings for models 1a (table 7) and 2a (table 11) suggests that the change in the model specification has very little impact on either the scores or the rankings. This is confirmed by table 12, which reports the correlation matrix for the scores from the DEA, COLS and SFA (half-normal) versions of models 1a and 2a. The correlation coefficient for the DEA scores from models 1a and 2a is 0.983, and there are similar coefficients for both the COLS ratings and the half-normal SFA scores.

11 Although the ratings for the truncated normal and exponential versions of the SFA model are perfectly correlated, their means and standard deviations do differ slightly (for example, the mean rating for the truncated normal model is 0.970409 while the mean rating for the exponential model is 0.970497).

Table 11: Individual PCT scores and rankings for model 2a

PCT	Efficiency scores (model 2a)					Efficiency rankings (model 2a)				
	DEA	COLS	SFA/HN	SFA/TN	SFA/E	DEA	COLS	SFA/HN	SFA/TN	SFA/E
A	0.906	0.906	0.973	0.981	0.982	64	108	106	106	106
B	0.997	0.994	0.989	0.991	0.991	140	151	151	151	151
C	0.943	0.874	0.960	0.974	0.974	108	57	60	62	62
D	0.934	0.912	0.975	0.983	0.983	97	118	117	116	116
E	0.926	0.916	0.976	0.984	0.984	89	125	124	128	128
F	0.953	0.878	0.963	0.976	0.976	115	63	65	69	71
G	1.000	0.963	0.986	0.989	0.989	152	147	148	148	148
H	1.000	0.925	0.979	0.985	0.985	152	135	135	135	135
I	1.000	0.938	0.981	0.987	0.987	152	144	143	143	143
J	0.957	0.902	0.973	0.982	0.982	121	105	107	107	107
K	0.879	0.888	0.963	0.975	0.975	33	85	69	65	65
L	0.978	0.975	0.987	0.990	0.990	133	149	149	149	149
M	0.903	0.898	0.969	0.979	0.979	58	100	94	91	91
N	0.792	0.765	0.874	0.871	0.871	2	1	1	1	1
O	0.931	0.931	0.980	0.986	0.986	95	140	138	138	138
P	0.943	0.897	0.971	0.981	0.981	109	98	101	102	102
Q	0.869	0.867	0.954	0.968	0.969	26	48	46	43	43
R	0.848	0.819	0.922	0.938	0.938	14	9	10	10	10
S	0.907	0.896	0.969	0.979	0.979	66	97	95	93	93
T	0.873	0.860	0.949	0.965	0.965	28	39	36	35	35

Note: The efficiency rankings are from 1 (=least efficient) to 152 (=most efficient).

Table 12: Correlation matrix for efficiency scores from models 1a and 2a

Model and estimation method	Model 1a			Model 2a		
	DEA	COLS	SFA/HN	DEA	COLS	SFA/HN
Model 1a: DEA	1					
Model 1a: COLS	0.749	1				
Model 1a: SFA/HN	0.774	0.947	1			
Model 2a: DEA	0.983	0.786	0.802	1		
Model 2a: COLS	0.759	0.977	0.932	0.794	1	
Model 2a: SFA/HN	0.792	0.926	0.982	0.817	0.947	1

Note: SFA/HN is the half-normal SFA model.

5.3 Results for model 3

Model 3 is similar to model 2, with a mortality indicator and the need for healthcare as outputs. However, the single input in model 2 – expenditure per head (adjusted for local input prices) – is now disaggregated into three parts: expenditure per head on cancer, expenditure per person on circulation problems, and expenditure per head on all other categories of healthcare. With multiple inputs and multiple outputs, we cannot readily estimate this model using an econometric approach but we can still obtain efficiency ratings from DEA and compare these ratings with those obtained for models 1 and 2.

Table 13 provides descriptive statistics for the DEA efficiency scores for the four variants of model 3. These statistics are almost identical for all four sets of scores and these scores are also very similar to those obtained when using DEA to estimate models 1 and 2 (see tables 3 and 8).

Table 13: Model 3 DEA technical efficiency scores

DEA scores	Model 3a	Model 3b	Model 3c	Model 3d
Mean	0.936	0.936	0.935	0.936
Std dev	0.048	0.049	0.047	0.049
Min	0.809	0.809	0.809	0.809
Max	1.000	1.000	1.000	1.000
Efficient PCTs	26	28	24	25

Table 14 reports the correlations between the efficiency scores for various models estimated using DEA (these all employ SMRs as the mortality indicator). The extremely high correlation (0.987) between the scores for models 3a and 3c confirms the idea, suggested by table 12, that the model 3 efficiency scores are very similar for all four variants. The correlation between model 3 efficiency ratings and those generated by both models 1 and 2 is slightly lower at about 0.90. There are very high correlations between models 1 and 2 (coefficients of about 0.98).

Table 14: Correlation matrix for DEA efficiency scores from models 1, 2 and 3

DEA model	Model 1a	Model 2a	Model 2c	Model 3a	Model 3c
Model 1a	1				
Model 2a	0.983	1			
Model 2c	0.977	0.990	1		
Model 3a	0.893	0.893	0.894	1	
Model 3c	0.877	0.876	0.896	0.987	1

6. Some detailed DEA results: three case studies

Hitherto we have focused on the efficiency ratings generated by various econometric and DEA approaches to efficiency measurement. However, DEA provides much more than a single efficiency rating for each unit. In this section we use the results from one DEA model – that with two outputs (the all-causes SMR and the need for healthcare) and one input (cost per person) – to illustrate the range of information provided by DEA.

6.1 Case study 1: PCT A

The DEA results for model 2a – with two outputs and one input – reveal a mean efficiency score of 0.915. PCT A records an efficiency rating of 0.9234. As we have specified an input orientation (because we believe that, in the context of this model, managers have more control over inputs than outputs), this implies that PCT A could reduce its expenditure per head by 7.66 per cent and still achieve its current output levels. In other words, DEA suggests that there is a weighted average of other PCTs which can be thought of as forming a ‘composite’ peer PCT, and that this composite produces at least as much of the two outputs as does PCT A but only uses 92.34 per cent of the input that PCT A uses.

Most DEA software will identify the peer group for each unit analysed. PCT A serves a relatively affluent population near London and its peers are the also affluent Berkshire West and Kensington PCTs. In addition to identifying the peer group, DEA software also reports the weights with which each DMU contributes to the composite peer. In this case, Berkshire West has a weight of 0.961 and Kensington has a weight of 0.031 – that is, PCT A is being compared with a composite PCT that has the characteristics of 96.1 per cent of Berkshire West and 3.1 per cent of Kensington.

Rows 1–3 in table 15 report the value of the two outputs and the one input for these three PCTs. If we multiply the Berkshire West data in row 2 by their weight we get the data in row 4, and if we multiply the Kensington data in row 3 by their weight we get the data in row 5. If we then sum the weighted data for Berkshire and Kensington (in rows 4 and 5), we obtain the composite peer (in row 6) with which PCT A is being compared.

The composite peer produces at least as much of both outputs as PCT A does but it does so at a lower cost (£996.2 rather than £1,078.9 per person). This cost is 92.34 per cent of PCT A’s cost and hence this is the efficiency rating recorded by PCT A. Note too that although the composite PCT and PCT A produce the same amount of output 2, the composite produces more of output 1 than does PCT A, and for this reason PCT A is said to exhibit slack on output 1.

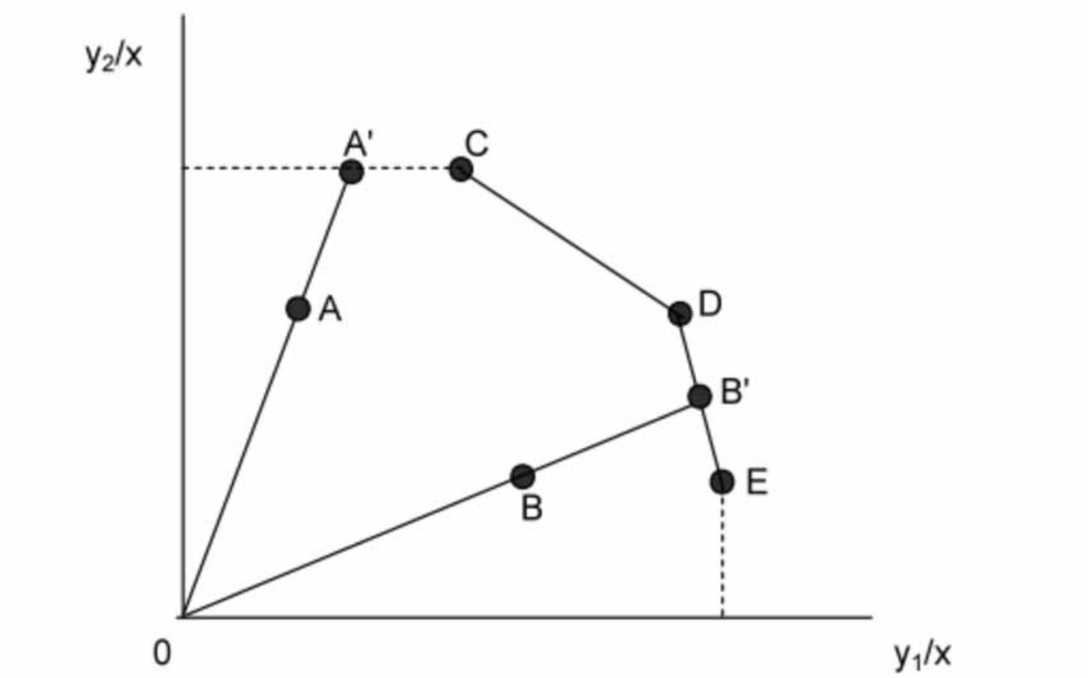
Table 15: Detailed DEA results for PCT A

PCT	Weight as peer	Output 1 (need)	Output 2 (SMR)	Input (cost per head)
1. PCT A		0.7581	0.003782	1078.9
2. Berkshire West		0.7680	0.003752	992.7
3. Kensington		0.7993	0.004723	1105.5
4. Berkshire West	0.969	0.744	0.003636	961.9
5. Kensington	0.031	0.025	0.000146	34.3
6. Berkshire and Kensington combined (PCT A's composite peer)		0.769	0.003782	996.2

Note: The SMR variable has been inverted to ensure that larger values reflect a better outcome.

To illustrate this result graphically, figure 9 depicts a situation with two outputs (y_1 and y_2) and a single input (x). DMUs A and B represent inefficient production units and C, D and E are efficient, forming the frontier. Thus, the inefficiency of units A and B are calculated as OA^1/OA and OB^1/OB respectively. A^1C represents the 'output slack' or the amount by which output y_1 can still be expanded without any additional input or loss of other output. PCT A could be at point A, with Kensington at point C and Berkshire West at point D.

Figure 9: Efficiency measurement and output slack



Whether point A^1 in figure 9 is efficient remains a moot issue in the literature. A stricter definition of efficiency would argue that DMUs are technically efficient only if they operate on the frontier (such as DMUs C and D) and all associated slacks are 0. Typically, efficiency measures are reported alongside non-0 input or output slacks to give a more accurate picture of efficiency,

6.2 Case study 2: PCT B

Detailed DEA results for PCT B are presented in table 16. PCT B records an efficiency rating of 0.997 so that it is almost 100 per cent efficient. An efficiency rating of 0.997 implies that PCT B could reduce its expenditure per head marginally (by 0.3 per cent) and still achieve its current output levels. PCT B's peer group comprises Redcar and Bedfordshire PCTs, with Redcar having a weight of 0.725 and Bedfordshire having a weight of 0.275.

Rows 1–3 in table 16 report the value of the two outputs and the one input for these three PCTs. If we multiply the Redcar data in row 2 by their weight we get the data in row 4, and if we multiply the Bedfordshire data in row 3 by their weight we obtain the data in row 5. If we then sum the weighted data for Redcar and Bedfordshire (in rows 4 and 5), we obtain the composite peer (in row 6) with which PCT B is being compared.

Table 16: Detailed DEA results for PCT B

PCT	Weight as peer	Output 1 (need)	Output 2 (SMR)	Input (cost per head)
1. PCT B		1.0980	0.002698	1276.6
2. Redcar		1.1885	0.002736	1356.7
3. Bedfordshire		0.8591	0.003479	1053.4
4. Redcar	0.725	0.8617	0.001956	983.6
5. Bedfordshire	0.275	0.2363	0.000956	289.7
6. Redcar and Bedfordshire combined (PCT B's composite peer)		1.0980	0.002912	1273.3

Note: The SMR variable has been inverted to ensure that larger values reflect a better outcome.

The composite peer produces at least as much of both outputs as PCT B does but it does so at a slightly lower cost (£1,273.3 rather than £1,276.6 per person). This cost is 99.7 per cent of PCT B's cost and hence this is the efficiency rating recorded by PCT B. Note too that although the composite PCT and PCT B produce the same amount of output 1, the composite produces more of output 2 than PCT B.

6.3 Case study 3: PCT C

Detailed DEA results for PCT C are presented in table 17. PCT C records an efficiency rating of 0.7919. This is one of the lower recorded ratings and implies that PCT C could reduce its expenditure per head by 20.81 per cent and still achieve its current output levels. PCT C's peer group comprises Knowsley and Redcar PCTs, with Knowsley having a weight of 0.483 and Redcar a weight of 0.517.

Rows 1–3 in table 17 report the value of the two outputs and the one input for these three PCTs. If we multiply the Knowsley data in row 2 by their weight we get the data in row 4, and if we multiply the Redcar data in row 3 by their weight we obtain the data in row 5. If we then sum the weighted data for Knowsley and Redcar (in rows 4 and 5), we obtain the composite peer (in row 6) with which PCT C is being compared.

Table 17: Detailed DEA results for PCT C

PCT	Weight as peer	Output 1 (need)	Output 2 (SMR)	Input (cost per head)
1. PCT C		1.2910	0.002205	1926.9
2. Knowsley		1.4006	0.002341	1707.1
3. Redcar		1.1885	0.002736	1356.7
4. Knowsley	0.483	0.6765	0.001131	824.5
5. Redcar	0.517	0.6145	0.001415	701.4
6. Knowsley and Redcar combined (PCT C's composite peer)		1.2910	0.002546	1525.9

Note: The SMR variable has been inverted to ensure that larger values reflect a better outcome.

The composite peer produces at least as much of both outputs as PCT C but it does so at a considerably lower cost (£1,525.9 rather than £1,926.9 per person). This cost is 79.19 per cent of PCT C's cost and hence this is the efficiency rating recorded by PCT C. Note too that although the composite PCT and PCT C produce the same amount of output 1, the composite produces more of output 2 than PCT C.

6.4 PCTs with a 100 per cent efficiency rating

PCTs that act as peers for other PCTs are, by definition, 100 per cent efficient. And any PCT which records the largest output per unit of input for any output will also be 100 per cent efficient, irrespective of its performance in terms of the other outputs. Together, these factors can lead to several DMUs being accredited with 100 per cent efficiency.

One way to differentiate between PCTs with a 100 per cent efficiency rating is to count the number of times each DMU acts as a peer for another DMU. Table 18 presents this information for model 2a. Redcar PCT acts as a peer for over 80 per cent of all PCTs so, in the context of model 2a, is a particularly influential PCT.

Table 18: Peer count for model 2a

PCT	Number of times each PCT is a peer for another PCT
Knowsley	18
Kensington	12
Bedfordshire	71
East and North Hertfordshire	51
Suffolk	38
Berkshire West	9
Berkshire East	5
Dorset	13
Redcar	124
Isle of Wight	5
Torbay	9

7. Sensitivity analysis

In this section we examine the sensitivity of the econometric efficiency ratings to the estimation of a more general cost function. The cost functions that we have estimated (see table 2 for details) are simplified versions of a Cobb-Douglas cost function (see equation 7). In particular, factor prices have been omitted from the regressor set, partly because the NHS operates a national pay scale and local variations have been incorporated through the appropriate adjustment of the dependent variable (that is, by adjusting the denominator of the dependent variable – the size of the population – for variations in local input prices). In addition to variables reflecting input prices, the more general translog cost function (see equation 8) also includes the squares and cross-products of the output terms. To examine the impact of these ten additional terms on the efficiency scores, the COLS and half-normal SFA regressions for model 1a were re-estimated and the corresponding efficiency scores calculated.[12] Descriptive statistics for these efficiency scores are presented in table 19, together with the efficiency ratings for the corresponding models without these additional regressors.

Table 19: Descriptive statistics for efficiency ratings derived from a basic and extended cost function for model 1a

Model 1a	COLS score	COLS score	SFA score	SFA score
	basic model	extended model	basic model	extended model
Mean	0.880	0.873	0.961	0.965
Std dev	0.039	0.037	0.020	0.016
Min	0.767	0.766	0.884	0.903
Max	1.000	1.000	0.990	0.990

It is clear from table 19 that the addition of these ten extra variables to the cost function has little effect on the average efficiency rating and its variance. Further confirmation of this is provided by table 20, which reports correlation coefficients for these efficiency ratings. The ratings from the basic and extended COLS cost function are highly correlated (with a coefficient of 0.953) and so too are the ratings from the basic and extended SFA half-normal cost function (with a coefficient of 0.963).

Table 20: Correlation coefficients for efficiency ratings derived from a basic and extended cost function for model 1a

Model 1a	COLS score	COLS score	SFA score	SFA score
	basic model	extended model	basic model	extended model
COLS basic	1.000			
COLS extended	0.953	1.000		
SFA basic	0.947	0.897	1.000	
SFA extended	0.917	0.943	0.963	1.000

12 There are four squared terms (for the three SMRs and the need index) and six cross-product terms. If all ten terms are included, all terms are insignificant. If a process of elimination of the least significant variable and re-estimation is followed, two of the additional terms are significant: the cancer and circulation mortality interaction term (with a negative sign) and the cancer and other mortality interaction term (with a positive sign).

8. Correlates of efficiency ratings

Having estimated various sets of efficiency ratings, analysts often conduct a second-stage analysis in which they attempt to identify those factors that determine estimated efficiency levels. Typically, a regression model is estimated, with the efficiency score as the dependent variable and a set of regressors that are thought to influence the efficiency rating. Because the DEA efficiency scores are right-censored at 1 (that is, no matter what the efficiency of the PCT, it cannot obtain a rating of more than 100 per cent), OLS regression is not appropriate for the estimation of this regression and, instead, a Tobit regression is estimated. This estimator allows for the right-censoring of the efficiency scores at 1.

As potential regressors, we are looking for those characteristics of the resources employed by a PCT that might affect its efficiency. Data on the characteristics of the capital stock are difficult to obtain but characteristics of (some of) the workforce are more readily obtainable. From the September 2006 General Medical Services (GMS) annual survey of GPs, we constructed three variables that might affect the efficiency of primary care:

a) the average age of all GPs in a PCT

b) the proportion of all GPs in a PCT who are female

c) the proportion of all GPs in a PCT whose medical qualification was obtained outside the UK.

Older GPs might be more efficient than their younger colleagues because they have more experience. We have no prior beliefs about the relative efficiency of female or non-UK trained GPs.[13]

As a further regressor we include the population weighted average of each PCT’s constituent local authority IMD2007 (deprivation) scores. For example, it is possible that PCTs in more affluent areas are able to recruit from a larger employee pool than those operating in more deprived areas and are thus able to select the most efficient staff.[14] If this is the case, affluence might be positively associated with efficiency because PCTs in more affluent areas find it easier to attract and retain more efficient staff than their counterparts operating in more deprived areas.[15]

Our expenditure data is for 152 PCTs and relates to the financial year 2006/07. At the beginning of this year there were 303 PCTs. In the middle of 2006/07, about two-thirds of all PCTs underwent a major reorganisation, with about 225 PCTs merged to form 75 new PCTs. The remaining PCTs – about 75 – were unaffected by these mergers. It is possible that these mergers were in part driven by efficiency considerations, and so we include a dummy variable in the model that takes a value of 1 if the PCT was formed by a merger of other PCTs in 2006/07.

We also include a variable that reflects the percentage by which each PCT’s actual budgetary allocation from the Department of Health exceeds its target allocation. Periodically, the formula employed by the Department of Health to determine each PCT's budgetary allocation is revised. This formula is designed to enable all PCTs to offer their populations the same standard of healthcare given local input prices, demographic and socio-economic conditions. To avoid sharp budgetary changes following the introduction of a new resource allocation formula, the Department of Health phases in the adjustment

13 It would be difficult to construct similar variables for secondary care because each PCT will buy secondary care from several hospital trusts.

14 More efficient staff might be those who make more effort and/or are more skilled than their colleagues.

15 We are indebted to our colleague Andy Street for this suggestion.

of a PCT's actual allocation to its new target allocation by spreading this adjustment over a number of years. It seems plausible that a PCT whose actual allocation exceeds its target allocation will be under less pressure to operate efficiently than a PCT whose actual budget falls short of its target budget.

We also tested for a scale effect by including the number of patients registered with a practice within the PCT.

We added several dummy variables to each regression to reflect the Audit Commission's ALE score and the Healthcare Commission's quality of services score awarded to each PCT for 2006/07. The Auditors' Local Evaluation (ALE) score reflects how well each PCT manages and uses its financial resources, with each PCT being awarded a rating of either 1 (=weak), 2 (=fair), 3 (=good) or 4 (=excellent) (Audit Commission, 2007). Similarly, the Healthcare Commission awards each PCT a score from 1 to 4, which reflects its assessment of the quality of services provided by the PCT (Healthcare Commission, 2007). We added six dummy variables to each regression for these ALE and Healthcare Commission ratings (three for the ALE score and three for the quality of services score, with a score of 1 (=weak performance) being the baseline).

We also added several dummy variables to reflect each PCT's 'World Class Commissioning' ratings (*Health Services Journal,* 5 March 2009). Each PCT was awarded a rating from 1 (=red) to 3 (=green) for three competencies: strategy, financial management and board skills. We added six dummy variables to each regression for these three ratings (two each for strategy, finance, and board skills), with a score of 1 being the baseline (and the worst rating) and 4 being the best rating but not achieved by any PCT for any competency. We also included the total competency score across all domains as an additional regressor.

Finally, we included the PCT population weighted average of each practice's QOF achievement score for eight disease areas in 2005. These scores reflect practices' performance with respect to 30 quality indicators covering clinical care for eight chronic diseases (see Doran et al (2006) for further details of these scores).

Table 21 reports descriptive statistics for the variables involved in the efficiency regressions. In total, we estimated seven models, and the dependent variables for these models are included in table 21. The average age of GPs across all 152 PCTs was just under 46 years but this varied from 41 years in Northumberland to just over 55 years in Barking and Dagenham. Just under 40 per cent of all GPs were female but this proportion varied from 22 per cent in North East Lincolnshire to 60 per cent in Richmond. Just over 26 per cent of GPs qualified outside the UK but this percentage varied from fewer than 5 per cent in Torbay to more than 77 per cent in Barking and Dagenham. For four PCTs, their actual budgetary allocation in 2006/07 was the same as their target allocation. Solihull's allocation in 2006/07 was almost 15 per cent above its target allocation, while eight PCTs received a budget that was 5.1 per cent below their target budget.

Table 21: Descriptive statistics for variables in efficiency regressions

Variable description	Variable name	Mean	Std Dev	Minimum	Maximum
Technical efficiency rating, DEA model 1a	DEATE 1a	0.922	0.051	0.776	1.000
Economic efficiency rating, COLS model 1a	TECOLS 1a	0.880	0.040	0.768	1.000
Economic efficiency rating, SFA-HN model 1a	TEHN 1a	0.961	0.020	0.885	0.990
Economic efficiency rating, DEA model 2a	DEATE 2a	0.916	0.050	0.776	1.000
Economic efficiency rating, COLS model 2a	TECOLS 2a	0.884	0.041	0.765	1.000
Economic efficiency rating, SFA-HN model 2a	TEHN 2a	0.959	0.022	0.874	0.990
Technical efficiency rating, DEA model 3a	DEATE 3a	0.936	0.048	0.810	1.000
GP age (years)	GPAGE	45.922	2.129	41.197	55.112
GP gender (=0 if female, =1 if male)	GPGENDER	0.396	0.078	0.218	0.605
GP qualified outside UK (0=no, 1=yes)	CQUALNUK	0.264	0.150	0.048	0.775
Index of Multiple Deprivation, 2007	IMD2007	23.633	9.068	8.063	46.970
New PCT dummy	NEWPCTDV	0.474	0.501	0.000	1.000
Distance from target (%, actual/target)	DFT67PC	0.145	4.185	–5.100	14.800
Patient list size	TOTPAT	349093	190496	92890	1265470
ALE use of financial resources score: dv2=1 if fair	ALE67DV2	0.513	0.501	0.000	1.000
ALE use of financial resources score: dv3=1 if good	ALE67DV3	0.164	0.372	0.000	1.000
ALE use of financial resources score: dv4=1 if excellent	ALE67DV4	0.033	0.179	0.000	1.000
HC quality of service: dv2=1 if fair	HCQ67DV2	0.618	0.487	0.000	1.000
HC quality of service: dv3=1 if good	HCQ67DV3	0.250	0.434	0.000	1.000
HC quality of service: dv4=1 if excellent	HCQ67DV4	0.013	0.114	0.000	1.000
Commissioning: total competency score	COMPETEN	16.546	2.460	11.000	23.000
Commissioning strategy score: dv2=1 if score is amber	STRATDV2	0.605	0.490	0.000	1.000
Commissioning strategy score: dv3=1 if score is green	STRATDV3	0.296	0.458	0.000	1.000
Commissioning finance score: dv2=1 if score is amber	FINANDV2	0.474	0.501	0.000	1.000
Commissioning finance score: dv3=1 if score is green	FINANDV3	0.428	0.496	0.000	1.000
Commissioning board score: dv2=1 if score is amber	BOARDDV2	0.533	0.501	0.000	1.000

Commissioning board score: dv3=1 if score is green	BOARDDV3	0.447	0.499	0.000	1.000
Asthma QOF achievement score, 2005	ASTHMA	73.925	2.288	68.150	78.913
CHD QOF achievement score, 2005	CHD	81.477	1.417	76.969	85.235
COPD QOF achievement score, 2005	COPD	83.634	2.693	76.375	90.206
Diabetes QOF achievement score, 2005	DIABETES	77.003	2.331	68.902	81.319
Hypertension QOF achievement score, 2005	BP	79.270	1.735	73.987	84.023
Thyroid function QOF achievement score, 2005	THYROID	95.393	1.230	90.948	97.737
Mental health QOF achievement score, 2005	MH	81.841	3.634	67.929	89.204
Stroke QOF achievement score, 2005	STROKE	79.539	1.603	74.024	83.841

Notes

1. The GP data are based on individual GP information and the figures for each PCT reflect these individual data weighted by each GP's full-time equivalent commitment.
2. The descriptive statistics across PCTs are unweighted.

The average PCT has just under 350,000 patients, with the smallest having fewer than 93,000 patients and the largest having more than 1.25 million. Just over half of all PCTs recorded a 'fair' rating for their use of resources in 2006/07, and over 60 per cent recorded a 'fair' rating for their quality of service in 2006/07. The average commissioning competency score was just over 16 and this ranged from 11 to 23. Just over 60 per cent of PCTs recorded an amber rating for their strategy but fewer than 50 per cent recorded an amber rating for their financial management.

Table 22 reports Tobit (for DEA efficiency scores) and OLS (for COLS/SFA scores) regressions for seven sets of efficiency ratings using the regressors listed in table 21.[16] [17] The first point to note is that the 'distance over target allocation' variable is significant at the 1 per cent level in all seven regressions and has the anticipated negative sign. It implies that efficiency declines as the actual allocation increases relative to the target allocation. One interpretation of this result is that PCTs with a budget that is relatively generous for the needs of its population (so that its actual budget exceeds its target budget) will be under less pressure to operate efficiently than those PCTs whose budget falls short of their target 'fair shares' allocation.

The second result to note from table 22 is that, in all three of the DEA regressions, efficiency is negatively associated with deprivation, so that PCTs operating in more deprived areas appear to be less efficient than their counterparts operating in less deprived areas. As noted above, if we assume that, on average, people prefer to live and work in affluent rather than deprived locations, it is possible that PCTs in more affluent areas are able to attract more efficient employees than PCTs operating in more deprived

16 The technical legitimacy of this type of exercise has been questioned by Simar and Wilson (2007). They argue that the DEA efficiency estimates are serially correlated because, in finite samples, perturbations of observations lying on the estimated frontier will in many cases cause changes in efficiencies estimated for other observations. Simar and Wilson also note that a similar, but less severe, problem arises with OLS regression. Although the coefficient estimates remain unbiased in the presence of this correlation, the variance of the OLS estimator is incorrect so that inference based on the reported standard errors may be misleading.

17 OLS estimation of the Tobit regressions generates very similar results.

areas. Another reason for this result may be that more affluent populations are better able to access NHS services. This might increase pressure on their PCTs to be more efficient than PCTs operating in a more deprived environment where patients are less demanding of their healthcare services. A further explanation may be that PCTs in deprived areas are devoting more resources to health inequality objectives, which are not included in our models. Finally, the inverse association between efficiency and deprivation may indicate that the ‘need for healthcare’ variable that we use in our models does not fully capture the impact of deprivation on need in the model that generates the efficiency rating. In this situation, PCTs in more deprived areas will appear to be operating in more favourable environmental conditions than they actually are and this will lead to the underestimation of their efficiency levels.

It is also worth noting that, in all three of the DEA regressions, the QOF COPD achievement score is significant at the 1 per cent level and has the anticipated positive sign. It is also significant in the four other equations at the 5 per cent level. However, and somewhat perversely, efficiency is negatively associated with the QOF achievement scores for coronary heart disease and diabetes in several of the equations. This may reflect the fact that the benefits of preventive activity are not immediately manifest in contemporary mortality rates.

The dummy variable for those PCTs awarded a ‘good’ ALE score was positive and significant in six of the seven regressions, and the dummy for an ‘excellent’ quality of service score was positive and significant in the three DEA regressions. There is also some evidence that the ‘board skills’ dummies have a positive impact on efficiency levels.

The use of Ramsey’s reset test revealed no evidence of misspecification in any of the seven regressions.

Table 22: Tobit/OLS model of efficiency scores when regressed against various PCT characteristics

Model and estimation technique generating the efficiency ratings																	
Regressors	Model 1a						Regressors	Model 2a						Regressors	Model 3a	
	DEA		COLS		SFA-HN			DEA		COLS		SFA-HN			DEA	
	Coeff	T-ratio	Coeff	T-ratio	Coeff	T-ratio		Coeff	T-ratio	Coeff	T-ratio	Coeff	T-ratio		Coeff	T-ratio
	1		2		3			4		5		6			7	
Constant	2.639	1.50	0.659	0.43	1.311	1.67	Constant	1.296	0.77	0.874	0.58	1.421	1.69	Constant	2.258	1.23
LGPAGE	0.147	1.15	0.037	0.34	0.002	0.03	LGPAGE	0.125	1.01	0.001	0.01	−0.010	−0.16	LGPAGE	0.230	1.70
LGPGENDE	0.021	0.74	0.008	0.34	-0.003	-0.27	LGPGENDE	0.018	0.66	0.003	0.14	−0.005	−0.34	LGPGENDE	0.033	1.09
LCQUALNU	0.001	0.16	0.007	0.93	0.003	0.72	LCQUALNU	−0.001	−0.18	0.010	1.33	0.004	0.95	LCQUALNU	0.003	0.35
LIMD2007	−0.067	−5.30	−0.008	−0.70	−0.006	−1.11	LIMD2007	−0.051	−4.19	−0.011	−0.99	−0.009	−1.51	LIMD2007	−0.041	−3.13
NEWPCTDV	0.022	2.30	0.016	1.97	0.007	1.68	NEWPCTDV	0.021	2.31	0.014	1.65	0.006	1.30	NEWPCTDV	0.020	2.00
DFT67PC	−0.006	−5.70	−0.005	−6.02	−0.003	−5.63	DFT67PC	−0.006	−5.86	−0.006	-6.45	-0.003	−5.79	DFT67PC	−0.005	−5.15
LTOTPAT	−0.012	−1.18	−0.008	−0.92	−0.002	−0.40	LTOTPAT	−0.011	−1.14	−0.006	-0.66	0.000	−0.11	LTOTPAT	−0.003	−0.32
ALE67DV2	0.008	0.93	0.004	0.55	0.002	0.56	ALE67DV2	0.002	0.27	0.005	0.66	0.003	0.72	ALE67DV2	0.007	0.72
ALE67DV3	0.029	2.34	0.024	2.18	0.014	2.56	ALE67DV3	0.023	1.94	0.026	2.46	0.016	2.73	ALE67DV3	0.030	2.28
ALE67DV4	0.046	2.28	0.030	1.76	0.013	1.47	ALE67DV4	0.035	1.83	0.034	2.00	0.015	1.58	ALE67DV4	0.033	1.53
HCQ67DV2	0.002	0.23	−0.006	−0.65	−0.001	−0.23	HCQ67DV2	0.001	0.11	−0.006	−0.72	−0.002	−0.35	HCQ67DV2	0.004	0.40
HCQ67DV3	0.001	0.08	−0.007	−0.62	−0.003	−0.52	HCQ67DV3	0.004	0.33	−0.005	−0.46	−0.003	−0.50	HCQ67DV3	0.015	1.07
HCQ67DV4	0.064	2.12	0.044	1.66	0.016	1.18	HCQ67DV4	0.065	2.21	0.049	1.87	0.018	1.27	HCQ67DV4	0.121	2.89
LCOMPET	0.001	0.04	0.003	0.13	0.002	0.12	LCOMPET	0.007	0.25	0.001	0.02	0.000	-0.02	LCOMPET	0.011	0.35
STRATDV2	0.002	0.18	0.005	0.46	0.004	0.67	STRATDV2	0.002	0.18	0.005	0.53	0.003	0.56	STRATDV2	0.001	0.11
STRATDV3	0.022	1.50	0.014	1.14	0.009	1.35	STRATDV3	0.019	1.36	0.014	1.11	0.008	1.19	STRATDV3	0.025	1.63
FINANDV2	−0.004	−0.35	−0.004	−0.35	−0.002	−0.43	FINANDV2	−0.001	−0.06	−0.006	−0.54	−0.003	−0.52	FINANDV2	−0.006	−0.42
FINANDV3	−0.016	−1.13	−0.011	−0.92	−0.007	−1.12	FINANDV3	−0.014	−1.02	−0.014	−1.15	−0.008	−1.23	FINANDV3	−0.017	−1.09
BOARDDV2	0.044	1.78	0.051	2.39	0.024	2.22	BOARDDV2	0.045	1.92	0.050	2.36	0.024	2.03	BOARDDV2	0.040	1.55
BOARDDV3	0.035	1.34	0.048	2.08	0.023	1.92	BOARDDV3	0.038	1.49	0.049	2.12	0.023	1.83	BOARDDV3	0.031	1.12
LASTHMA	0.237	1.69	0.139	1.14	0.098	1.55	LASTHMA	0.238	1.75	0.141	1.17	0.098	1.46	LASTHMA	0.292	1.97
LCHD	−0.743	−1.76	−0.255	−0.70	−0.123	−0.65	LCHD	−0.777	−1.92	−0.238	−0.66	−0.119	−0.59	LCHD	−0.978	−2.18
LCOPD	0.510	3.21	0.333	2.41	0.157	2.20	LCOPD	0.528	3.44	0.319	2.33	0.154	2.03	LCOPD	0.521	3.13
LDIABETE	−0.384	−2.25	−0.458	−3.08	−0.135	−1.75	LDIABETE	−0.425	−2.58	−0.489	−3.31	−0.161	−1.96	LDIABETE	−0.193	−1.06
LBP	0.345	1.31	0.301	1.31	0.126	1.06	LBP	0.404	1.59	0.344	1.52	0.162	1.29	LBP	0.156	0.56
LTHYROID	−0.272	−0.69	0.145	0.42	−0.080	−0.45	LTHYROID	0.072	0.19	0.102	0.30	−0.102	−0.54	LTHYROID	−0.262	−0.63
LMH	−0.067	−0.81	0.014	0.20	−0.013	−0.37	LMH	−0.033	−0.41	0.025	0.35	−0.010	−0.26	LMH	0.013	0.15
LSTROKE	−0.064	−0.20	−0.193	−0.68	−0.106	−0.73	LSTROKE	−0.155	−0.50	−0.196	−0.70	−0.115	−0.74	LSTROKE	−0.013	−0.04
Adjusted R^2	0.4468		0.3571		0.3247		Adjusted R^2	0.4297		0.3980		0.3347		Adjusted R^2	0.3136	

Notes: see page 60

1. The dependent variable is the efficiency rating and, with the exception of the distance from target variable and the dummy variables, all other regressors are in natural logarithms (but not the dependent variable). The distance from target variable is not logged because, for many PCTs, its value is negative (that is, its actual allocation is less than its target allocation).
2. The SFA efficiency ratings are from the half-normal model.
3. The models with DEA scores as the dependent variable use a Tobit estimator while the other models use OLS.
4. The adjusted R-squared for the DEA models are from the estimation of the same model but using OLS (which generates very similar coefficients to the Tobit estimator).
5. Each model was re-estimated with the addition of the square of the predicted value (that is, Ramsey's reset test was undertaken). This is a general test of model specification. This additional variable was insignificant in all seven models.
6. Variables are defined in table 21.

The inclusion of a relatively large number of variables in our model might make it difficult to identify significant regressors if the variables are highly correlated with each other (see section A2 in the appendix for a correlation matrix of the variables in our regression models). Through a repeated process of estimation, dropping the least significant regressor and re-estimation, we were able to identify models where the only remaining regressors were significant at the 5 per cent level. The results from the application of this process to models 1–3 in table 22 are provided in table 23. These results largely confirm our previous findings. The 'distance from target allocation' variable is highly significant in all three models, as is the COPD achievement score. There is also evidence of the previously observed negative association between deprivation and efficiency, and the positive association between the use of resources/quality of service indicators and efficiency. There is also evidence of the previously observed and expected positive association between board skills and efficiency, as well as the previously observed but unanticipated negative association between efficiency and the QOF-based diabetes achievement score.

Table 23: Parsimonious Tobit/OLS models of efficiency scores from model 1a when regressed against various PCT characteristics

Dependent variable is efficiency rating from the application of model 1a

Regressors	DEA		Regressors	COLS		Regressors	SFA-HN	
	Coeff 1	T-ratio		Coeff 2	T-ratio		Coeff 3	T-ratio
Constant	2.950	3.115	Constant	1.609	2.327	Constant	1.179	2.930
LCQUALNU			LCQUALNU			LCQUALNU	0.007	2.710
LIMD2007	–0.066	–6.696	LIMD2007			LIMD2007	–0.011	–2.634
DFT67PC	–0.006	–6.683	DFT67PC	–0.005	–7.997	DFT67PC	–0.003	–7.624
ALE67DV3			ALE67DV3			ALE67DV3	0.010	2.634
HCQ67DV4	0.069	2.355	HCQ67DV4			HCQ67DV4		
BOARDDV2			BOARDDV2	0.057	2.948	BOARDDV2	0.024	2.409
BOARDDV3			BOARDDV3	0.058	2.962	BOARDDV3	0.023	2.299
LASTHMA	0.361	2.901	LASTHMA			LASTHMA		
LCHD	–0.808	–2.924	LCHD	–0.496	–2.079	LCHD	–0.232	–2.294
LCOPD	0.371	2.405	LCOPD	0.348	3.280	LCOPD	0.186	3.162
LDIABETE	–0.337	–2.276	LDIABETE	–0.509	–4.149	LDIABETE		
LBP			LBP	0.474	2.791	LBP		
Adjusted R^2	0.439		Adjusted R^2	0.351		Adjusted R^2	0.328	

Note: See table 22.

9. Conclusions

In England in 2006/07, 152 PCTs were responsible for about £66bn of the publicly funded healthcare budget. We employed the two major tools of efficiency analysis – data envelopment analysis (DEA) and econometric analysis (COLS/SFA) – to examine the efficiency of this expenditure in terms of reducing mortality rates given the local need for healthcare. We found that the average efficiency score was about 0.88 for COLS, about 0.92 for DEA and 0.96 for SFA. These scores suggest that, although there may be some scope for improvement, the potential efficiency gains are unlikely to be substantial. The three SFA variants yielded very similar efficiency scores. Given a particular estimation technique, the results are rather insensitive to model specification. The SFA efficiency ratings are highly correlated with the COLS ratings (with a correlation coefficient of about 0.90–0.95) but are less highly correlated with the DEA ratings (with a coefficient of about 0.75–0.80).

We also regressed the efficiency score on the characteristics of the resources employed by the PCT and identified several factors associated with efficiency: first, ‘underfunding’ is associated with increased efficiency; second, deprivation is associated with reduced efficiency, even after taking account of relative needs; and third, the QOF-based COPD achievement score is positively associated with efficiency. There is also some evidence that the ratings awarded to PCTs for the use of financial resources, the quality of services provided and the quality of commissioning are also positively associated with efficiency levels.

Of course we recognise that this preliminary study has many limitations. It uses limited health outcomes data (in the form of mortality rates). PCTs buy a wide range of services for their population and some are directed to improving the health-related quality of life rather than prolonging it. As we employ mortality rates as our only outcome/output indicator, our results may be biased against those PCTs that perform well in the provision of services that improve the quality of life. Also, a specific output that we have not been able to measure is the PCTs’ success in addressing health inequalities. In recent years, this has been an important policy objective. However, incorporating equity into efficiency measurement is not at all straightforward. One of the reasons we find that PCTs in deprived areas appear to be less efficient may be that they are devoting more resources to equity objectives.

Furthermore, we have modelled outcome data for the three-year period 2004/06 along with expenditure data for the single year 2006/07. In practice, health outcomes are the result of years of expenditure by local PCTs and, conversely, current expenditure is expected to yield outcome benefits beyond the current year. Implicitly, our analysis assumes that PCTs have reached some sort of equilibrium in the expenditure choices they make and the outcomes they secure. This is probably not an unreasonable assumption given the relatively slow pace at which both types of variable change. But a longer time series of data would enable us to model the effects with more confidence.

Notwithstanding these limitations, our results are reasonably consistent across a variety of model specifications and estimation techniques. They illustrate the sort of analysis that is possible using DEA and econometric approaches to the estimation of efficiency, and they offer another perspective on the debate about efficiency and the quality of commissioning in the healthcare sector. From a policy perspective, it is essential that such an analysis is undertaken as part of the regulatory process to assure that the NHS is maintaining satisfactory levels of value for money. Although we found only a modest element of relative inefficiency, there were a small number of outlier PCTs that appear to be achieving significantly lower levels of efficiency than their otherwise identical counterparts, and there is therefore a strong case for further scrutiny of such PCTs. As more extensive and longer time series of data become available, this sort of analysis will yield increasingly useful insights into PCT performance.

Finally, it is worth underlining the fact that this analysis examines relative efficiency among PCTs, and therefore says nothing about the overall efficiency with which the NHS commissions health. There may be substantial system-wide initiatives that could improve efficiency across the board, and this study should not lead policy makers to the conclusion that further efficiencies cannot be secured. However, with a few exceptions, seeking to pick out underperforming individual PCTs would not appear to be especially fruitful.

References

Anand S, Ammar W, Evans T, Hasegawa T, Kissimova-Skarbek K, Langer A, Lucas AO et al (2002). *Report of the Scientific Peer Review Group on health systems performance assessment.* Geneva: World Health Organization.

Appleby J, Crawford R and Emmerson C (2009). *How cold will it be? Prospects for NHS funding: 2011–17.* London: The King's Fund.

Audit Commission (2007). *Review of the NHS financial year 2006/07.* London: HMSO.

Buck D (2000). 'The efficiency of community dental services in England: a data envelopment analysis'. *Community Dentistry and Oral Epidemiology,* 28, pp 274–280.

Christensen LR and Greene WH (1976). 'Economies of scale in US electric power generation'. *Journal of Political Economy,* 84, pp 655–676.

Christensen LR, Jorgenson DW and Lau LJ (1973). 'Transcendental logarithmic production functions'. *Review of Economics and Statistics,* 55, pp 28–45.

Department of Health (2005). *Unified exposition book: 2003/04, 2004/05 and 2005/06 PCT revenue resource limits.* London: DH.

Department of Health (2007a). *Gershon Efficiency Programme 2004–2008. Efficiency technical note.* London: DH.

Department of Health (2007b). *Payment by results: tariff information.* London: DH. See www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/DH_077279.

Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U and Roland M (2006). 'Pay for performance programs in family practices in the United Kingdom'. *New England Journal of Medicine,* 355, pp 375–384.

Farrell MJ (1957). 'The measurement of productive efficiency'. *Journal of the Royal Statistical Society,* Series A, 120, pp 253–290.

Gershon P (2004). *Releasing resources for the front line: independent review of public sector efficiency.* London: HM Treasury.

Giuffrida A (1999). 'Productivity and efficiency changes in primary care: a Malmquist index approach'. *Healthcare Management Science,* 2, pp 11–26.

Giuffrida A and Gravelle H (2001). 'Measuring performance in primary care: econometric analysis and DEA'. *Applied Economics,* 33, pp 163–175.

Giuffrida A, Gravelle H and Sutton M (2000). 'Efficiency and administrative costs in primary care'. *Journal of Health Economics,* 19, pp 983–1006.

Healthcare Commission (2007). *The annual health check 2006/07.* London: HMSO.

Hollingsworth B (2003). 'Non-parametric and parametric applications measuring efficiency in healthcare'. *Healthcare Management Science,* 6, pp 203–218.

Hollingsworth B and Parkin D (1995). 'The efficiency of Scottish acute hospitals: an application of data envelopment analysis'. *IMA Journal of Mathematics Applied in Medicine & Biology,* 12, pp 161–173.

Hollingsworth B and Peacock S (2008). *Efficiency measurement in health and heath care.* London: Routledge.

Hollingsworth B and Smith PC (2003). ‘The use of ratios in data envelopment analysis’. *Applied Economics Letters,* 10, pp 733–735.

Intriligator MD (1978). *Econometric models, techniques and applications.* Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Jacobs R (2001). ‘Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis’. *Healthcare Management Science,* 4, pp 103–116.

Jacobs R, Smith PC and Street A (2006). *Measuring efficiency in healthcare.* Cambridge University Press.

Johnston K and Gerard K (2001). ‘Assessing efficiency in the UK breast screening programme’. *Health Policy,* 56, pp 21–32.

Kerr C, Glass J, McCallion G and McKillop D (1999). ‘Best practice measures of resource utilisation for hospitals’. *Public Administration,* 77, pp 639–650.

Lakhani A, Olearnik H and Eayres D eds (2006). *Compendium of clinical and health indicators: data definitions and user guide for computer files.* London: NCHOD.

Maniadakis N and Thanassoulis E (2000). ‘Assessing productivity changes in UK hospitals reflecting technology and input prices’. *Applied Economics,* 32, pp 1575–1589.

Martin S, Rice N and Smith PC (2008). ‘Does healthcare spending improve health outcomes? Evidence from English programme budgeting data’. *Journal of Health Economics,* 27(4), pp 826–842.

Office for National Statistics (2008). *Public service productivity: healthcare.* London: ONS.

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, pp 1425–1433.

Salinas-Jiménez J and Smith PC (1996). ‘Data envelopment analysis applied to quality in primary healthcare’. *Annals of Operations Research,* 67, pp 141–161.

Simar L and Wilson PW (2007). ‘Estimation and inference in two-stage semi-parametric models of production processes’. *Journal of Econometrics,* 136, pp 31–64.

Smith PC (1997). ‘Model misspecification in data envelopment analysis’. *Annals of Operations Research,* 73, pp 233–252.

Smith PC (2002). *Measuring up: improving health system performance in OECD countries.* Paris: OECD.

Smith PC, Rice N and Carr-Hill RA (2001). ‘Capitation funding in the public sector’. *Journal of the Royal Statistical Society,* Series A, 164, Part 2, pp 217–257.

Street A (2003). ‘How much confidence should we place in efficiency estimates?’ *Health Economics,* 12, pp 895–907.

Thanassoulis E (2001). *Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software.* Dordrecht: Kluwer Academic Publishers.

Wanless D (2002). *Securing our future health: taking a long-term view.* London: HM Treasury.

Williams A (2001). ‘Science or marketing at WHO? A commentary on “World Health 2000”’. *Health Economics,* 10, pp 93–100.

World Health Organization (2000). *World health report 2000.* Geneva: WHO.

Appendix

A1 Regressions generating the COLS efficiency scores

The COLS efficiency scores reported in section 5 have been generated by the OLS regressions in table A1. We do not report the regressions generating the SFA efficiency scores because these are very similar to the COLS results.

Table A1: Regressions generating the COLS efficiency scores

Model	1a	1b	2a	2b	2c	2d
Constant	6.812** (0.234)	6.765** (0.282)	6.871** (0.207)	6.747** (0.246)	7.145** (0.142)	7.176** (0.155)
Cancer SMR	–0.042 (0.063)					
Circulation SMR	0.100* (0.046)					
Other SMR	–0.130** (0.047)					
Need for care	0.839** (0.045)	0.824** (0.052)	0.855** (0.043)	0.833** (0.050)	0.899** (0.043)	0.906** (0.045)
Cancer SYLL rate		–0.013 (0.063)				
Circulation SYLL rate		0.034 (0.034)				
Other SYLL rate		–0.095* (0.043)				
All-causes SMR			–0.054 (0.036)			
All-causes SYLL rate				–0.070 (0.040)		
Amenable SMR					–0.007 (0.030)	
Amenable SYLL rate						–0.001 (0.031)
Adj R-squared	0.876	0.874	0.872	0.873	0.870	0.870
Ramsey's test (F)	0.218	0.671	0.107	0.239	0.225	0.245

Notes

1. The dependent variable in all regressions is the logarithm of (total PCT expenditure per person).

2. For comparability with the DEA models, the mortality indicators have been inverted.
3. All regressors are measured in natural logarithms.
4. Coefficients are reported with standard errors in parentheses.
5. Ramsey's test is a general test of model specification. It is implemented by adding the square of the predicted value to the regression and re-estimating the model. The test statistic has an F-distribution. The test statistics for all six models are insignificant.

These OLS results are a little disappointing because only the need for healthcare appears to have a consistent and significant effect on expenditure. In part, this will reflect how we have measured need and the fact that this need index is part of the resource allocation formula that determines each PCT's target budget. Moreover, the significant negative coefficient on the 'other diseases' survival rate (remember that we use the reciprocal of the SMR) in model 1a may reflect the fact that the all-condition need variable does not fully capture the need for care in this disease category. Consequently, this negative coefficient may partly reflect the fact that areas with a high survival rate for the 'other disease' category also have low need and so attract less expenditure (rather than it being the case that increased survival reduces costs).

The results in table A1 are likely to be difficult to interpret because the relationship between expenditure, need and mortality is likely to be highly complex, involving some element of simultaneity between the variables. To build a more satisfactory model would probably require a system of equations and estimation via more advanced econometric techniques. In these circumstances one advantage associated with DEA is that it is not necessary to specify a comprehensive model and that, as a result, it can sidestep problems of simultaneity (Salinas-Jiménez and Smith, 1996). Despite these problems, the COLS/SFA efficiency ratings are highly correlated with those generated by DEA and similar factors seem to be associated with both the DEA and COLS/SFA efficiency scores.

Table A2 shows the correlations between the need, expenditure and mortality rate variables in the COLS regressions. Need and expenditure per head are very highly correlated (with a coefficient of 0.934). All the mortality rates are highly correlated with each other (with coefficients over 0.80) and these rates are strongly correlated with expenditure per head. The negative correlation coefficients imply that expenditure and mortality are positively correlated (remember that the mortality rates have been inverted). This positive correlation is probably detecting a budgetary effect: areas with high need also have high mortality rates and so attract larger budgets and have higher expenditure levels. We are more interested in the reverse process whereby higher expenditure levels lead to lower mortality rates given the local need for healthcare. This suggests that the mortality rates may be endogenous. The implications of this for our modelling will be considered in later work.

Table A2: Correlations between the need, expenditure, and mortality rate variables in the COLS regressions

	Spend per person	Need	Cancer SMR	Circulation SMR	Other SMR	All-causes SMR	Amenable causes SMR
Spend per person	1.000						
Need	.934	1.000					
Cancer SMR	-.734	-.762	1.000				
Circulation SMR	-.688	-.725	.836	1.000			
Other SMR	-.749	-.759	.829	.915	1.000		
All-causes SMR	-.758	-.782	.914	.966	.972	1.000	
Amenable causes SMR	-.708	-.753	.858	.981	.930	.972	1.000

Note: The SMRs are inverted.

Table A3 provides the correlation matrix for the variables employed in the regression models used to examine the correlates of the efficiency ratings (see section 8).

Table A3: Correlation matrix for variables used in the regression models of section 8

	DEATE1I	TECOLS1	TEHN1	DEATE3I	TECOLS3	TEHN3	DEATE7I	LGPAGE
DEATE1I	1.00000	.74979	.77449	.98347	.75939	.79291	.89339	.06528
TECOLS1	.74979	1.00000	.94706	.78636	.97736	.92612	.74706	.21769
TEHN1	.77449	.94706	1.00000	.80265	.93221	.98264	.77024	.16383
DEATE3I	.98347	.78636	.80265	1.00000	.79493	.81702	.89311	.05041
TECOLS3	.75939	.97736	.93221	.79493	1.00000	.94796	.75183	.23683
TEHN3	.79291	.92612	.98264	.81702	.94796	1.00000	.77617	.17449
DEATE7I	.89339	.74706	.77024	.89311	.75183	.77617	1.00000	.12359
LGPAGE	.06528	.21769	.16383	.05041	.23683	.17449	.12359	1.00000

	DEATE1I	TECOLS1	TEHN1	DEATE3I	TECOLS3	TEHN3	DEATE7I	LGPAGE
LGPGENDE	-.09078	-.30174	-.28887	-.10431	-.32267	-.29099	-.08269	-.52900
LCQUALNU	-.09392	.15378	.07558	-.09986	.19409	.09614	.02094	.70296
LIMD2007	-.43582	-.01084	-.08691	-.39852	-.00475	-.10507	-.22303	.32644
NEWPCTDV	.19681	.11232	.16000	.20163	.08439	.13950	.10943	-.19365
DFT67PC	-.29644	-.47401	-.46695	-.32560	-.49312	-.47043	-.27270	-.11937
LTOTPAT	.14176	.00305	.06625	.13566	-.00512	.07244	.08370	-.19174
ALE67DV2	-.13904	-.09691	-.10539	-.13949	-.10516	-.11323	-.11710	-.13640
ALE67DV3	-.02856	.10049	.10664	-.02362	.11131	.10659	.08398	.16611

	LGPGENDE	LCQUALNU	LIMD2007	NEWPCTDV	DFT67PC	LTOTPAT	ALE67DV2	ALE67DV3
LGPGENDE	1.00000	-.39828	-.25005	-.13597	.58009	.19378	-.12324	.02954
LCQUALNU	-.39828	1.00000	.47730	-.31372	-.04647	-.33918	-.04831	.12677
LIMD2007	-.25005	.47730	1.00000	-.32272	-.11370	-.39362	.15993	.28455
NEWPCTDV	-.13597	-.31372	-.32272	1.00000	-.20729	.61804	.10683	-.31428
DFT67PC	.58009	-.04647	-.11370	-.20729	1.00000	-.00868	.01045	.06842
LTOTPAT	.19378	-.33918	-.39362	.61804	-.00868	1.00000	-.11193	-.19689
ALE67DV2	-.12324	-.04831	.15993	.10683	.01045	-.11193	1.00000	-.45551
ALE67DV3	.02954	.12677	.28455	-.31428	.06842	-.19689	-.45551	1.00000

	DEATE1I	TECOLS1	TEHN1	DEATE3I	TECOLS3	TEHN3	DEATE7I	LGPAGE
ALE67DV4	.03124	.04537	.01475	.01731	.05980	.02354	-.00378	.13431
HCQ67DV2	.06112	-.02453	.00842	.04391	-.04873	-.00554	-.02962	.08441
HCQ67DV3	-.18072	-.04268	-.07553	-.15522	-.01410	-.06830	-.03576	.01005
HCQ67DV4	.06949	.12676	.08877	.07577	.13418	.08974	.12490	-.07374
LCOMPET	-.03362	.09872	.09591	-.01425	.08496	.07474	.04213	-.08771
STRATDV2	.01195	-.02628	-.01419	-.00028	-.01214	-.00618	-.03192	.15438
STRATDV3	-.01480	.08584	.07321	-.00306	.07607	.05786	.07479	-.11066
FINANDV2	.17780	.01568	.02691	.15032	.01833	.03943	.12722	.08676

	LGPGENDE	LCQUALNU	LIMD2007	NEWPCTDV	DFT67PC	LTOTPAT	ALE67DV2	ALE67DV3
ALE67DV4	-.05944	.11420	.05153	-.17496	-.00728	-.07126	-.18935	-.08183
HCQ67DV2	-.02661	-.05491	-.16701	.12134	.04934	.11775	-.14190	-.05336
HCQ67DV3	-.02228	.20737	.37574	-.30429	-.03060	-.34253	.19758	.15369
HCQ67DV4	-.02154	.03712	.13261	.00609	-.07598	-.04650	.11247	-.05123
LCOMPET	-.05236	-.09283	.21454	.12175	-.14106	.11767	.09941	.18029
STRATDV2	-.01478	.11613	-.13567	.01135	.13709	.02841	-.05953	-.00478
STRATDV3	-.01820	-.01930	.24061	-.06684	-.12374	-.08967	.08384	.10102
FINANDV2	.16178	.06302	-.26737	-.16111	.19105	-.02268	-.02497	-.10102

	ALE67DV4	HCQ67DV2	HCQ67DV3	HCQ67DV4	LCOMPET	STRATDV2	STRATDV3	FINANDV2
ALE67DV4	1.00000	.06894	-.02130	-.02130	.10008	-.07745	.12279	-.02722
HCQ67DV2	.06894	1.00000	-.73500	-.14700	-.06256	.05833	-.17292	.01285
HCQ67DV3	-.02130	-.73500	1.00000	-.06667	.06680	-.06217	.19137	-.12172
HCQ67DV4	-.02130	-.14700	-.06667	1.00000	.04349	-.02487	.05159	.00609
LCOMPET	.10008	-.06256	.06680	.04349	1.00000	-.30469	.54431	-.28201
STRATDV2	-.07745	.05833	-.06217	-.02487	-.30469	1.00000	-.80303	.22700
STRATDV3	.12279	-.17292	.19137	.05159	.54431	-.80303	1.00000	-.24001
FINANDV2	-.02722	.01285	-.12172	.00609	-.28201	.22700	-.24001	1.00000

	DEATE1I	TECOLS1	TEHN1	DEATE3I	TECOLS3	TEHN3	DEATE7I	LGPAGE
FINANDV3	-.24648	-.02164	-.05220	-.21896	-.03081	-.07132	-.17091	-.08636
BOARDDV2	.10504	-.04158	-.02385	.08691	-.04480	-.02613	.07590	-.00465
BOARDDV3	-.09422	.07064	.05535	-.07523	.07007	.05148	-.05907	-.01286
LASTHMA	.23895	.13442	.15965	.24522	.13513	.15726	.22908	-.03428
LCHD	.06845	-.00975	.02664	.08226	-.01068	.02683	-.02378	-.42102
LCOPD	.31887	.08666	.13671	.33426	.06781	.12407	.22664	-.43746
LDIABETE	.00454	-.09753	-.01237	.01729	-.11342	-.03118	-.03482	-.43647
LBP	.18683	.11357	.12368	.20113	.12719	.13802	.10175	-.07110

	LGPGENDE	LCQUALNU	LIMD2007	NEWPCTDV	DFT67PC	LTOTPAT	ALE67DV2	ALE67DV3
FINANDV3	-.16999	-.03272	.38955	.05887	-.17218	-.12058	.15018	.19045
BOARDDV2	.25821	.03368	-.22422	-.11537	.25537	.04378	-.09407	-.08261
BOARDDV3	-.23948	-.05265	.22742	.12692	-.23900	-.01773	.10867	.10050
LASTHMA	.11256	-.02500	-.21087	.07483	.10702	.09275	.07468	-.05442
LCHD	.02309	-.35623	-.34350	.28865	-.19232	.22163	.12827	-.31555
LCOPD	.17050	-.51345	-.50328	.21667	-.01641	.14329	.09625	-.27699
LDIABETE	-.06878	-.41795	-.22421	.34631	-.21734	.11244	.15354	-.27344
LBP	-.10180	-.05306	-.27672	.18449	-.17883	.08255	.06970	-.26061

	ALE67DV4	HCQ67DV2	HCQ67DV3	HCQ67DV4	LCOMPET	STRATDV2	STRATDV3	FINANDV2
FINANDV3	.06426	-.08753	.17658	.01689	.47564	-.25416	.37159	-.82001
BOARDDV2	-.12306	.07894	-.03807	-.00761	-.51420	.32302	-.43271	.41282
BOARDDV3	.13080	-.08315	.03056	.01222	.58439	-.30204	.45992	-.40307
LASTHMA	-.10095	.03569	-.01026	-.08976	-.06422	.03266	-.03771	-.00142
LCHD	-.02269	-.03031	-.00572	.08230	.03449	-.10279	.11325	-.13699
LCOPD	-.14005	-.00643	-.07402	.02756	-.08851	-.00893	-.01699	-.01563
LDIABETE	.00116	-.07775	.00734	.09008	.05628	-.15080	.14616	-.22141
LBP	.03405	-.05078	.02011	.03143	-.08352	-.02221	.02774	-.11266

	FINANDV3	BOARDDV2	BOARDDV3	LASTHMA	LCHD	LCOPD	LDIABETE	LBP
FINANDV3	1.00000	-.52343	.55952	-.05966	.10934	-.07337	.18802	.06452
BOARDDV2	-.52343	1.00000	-.96101	.13300	-.03759	.06214	-.16821	-.04966
BOARDDV3	.55952	-.96101	1.00000	-.13583	.05517	-.05993	.18914	.02482
LASTHMA	-.05966	.13300	-.13583	1.00000	.32319	.48346	.19003	.48559
LCHD	.10934	-.03759	.05517	.32319	1.00000	.59178	.65140	.66973
LCOPD	-.07337	.06214	-.05993	.48346	.59178	1.00000	.52670	.49304
LDIABETE	.18802	-.16821	.18914	.19003	.65140	.52670	1.00000	.57065
LBP	.06452	-.04966	.02482	.48559	.66973	.49304	.57065	1.00000

	DEATE1I	TECOLS1	TEHN1	DEATE3I	TECOLS3	TEHN3	DEATE7I	LGPAGE
LTHYROID	.06027	.06665	.06814	.10921	.05248	.05499	-.03722	-.35913
LMH	.03578	.09616	.04891	.05693	.11290	.06247	.06221	.19157
LSTROKE	-.01884	-.05605	-.03447	-.01377	-.05347	-.03591	-.04563	-.29043

	LGPGENDE	LCQUALNU	LIMD2007	NEWPCTDV	DFT67PC	LTOTPAT	ALE67DV2	ALE67DV3
LTHYROID	-.19853	-.33136	-.20482	.33275	-.28602	.07540	.04447	-.19510
LMH	-.05347	.21503	-.02005	-.04564	.09616	-.02289	-.01017	-.03162
LSTROKE	.09393	-.21368	-.11431	.04172	-.06906	.04965	.08526	-.13564

	ALE67DV4	HCQ67DV2	HCQ67DV3	HCQ67DV4	LCOMPET	STRATDV2	STRATDV3	FINANDV2
LTHYROID	.02932	.08611	-.12329	.03836	.10324	-.21222	.14884	-.30500
LMH	-.03706	.01919	.00874	-.17097	-.10001	.08987	-.03752	.05651
LSTROKE	.12093	-.06950	.08793	.11875	.10629	-.11751	.19617	-.12931

	FINANDV3	BOARDDV2	BOARDDV3	LASTHMA	LCHD	LCOPD	LDIABETE	LBP
LTHYROID	.25655	-.23215	.21855	.19786	.62882	.46298	.61633	.51504
LMH	-.02138	-.00411	.03086	.39652	.01013	.06619	-.11138	.22767
LSTROKE	.16437	-.05474	.08009	.34514	.80317	.45074	.57983	.66400

	LTHYROID	LMH	LSTROKE
LTHYROID	1.00000	-.05289	.44812
LMH	-.05289	1.00000	.03437
LSTROKE	.44812	.03437	1.00000

Note: For definitions of variable names see table 21. Where the natural logarithm of the variables is employed, the variable name is given the prefix ‘L’.

The Health Foundation
90 Long Acre
London
WC2E 9RA
T 020 7257 8000
F 020 7257 8001
E info@health.org.uk
Registered charity number 286967
Registered company number 1714937

www.health.org.uk