CENTRE FOR HEALTH ECONOMICS
THE UNIVERSITY of York

National Institute
of Economic and
Social Research

Developing New Approaches to Measuring NHS Outputs and Productivity

Lead Investigators:

Diane Dawson
Hugh Gravelle
Paul Kind
Mary O'Mahony
Andrew Street
Martin Weale

With the assistance of:

Adriana Castelli
Rowena Jacobs
Sue Macran
Casey Quinn
Philip Stevens
Lucy Stokes

CHE Technical Paper Series 31

CENTRE FOR HEALTH ECONOMICS TECHNICAL PAPER SERIES

The Centre for Health Economics has a well established Discussion Paper series which was originally conceived as a means of circulating ideas for discussion and debate amongst a wide readership that included health economists as well as those working within the NHS and pharmaceutical industry.

The introduction of a Technical Paper Series offers a further means by which the Centre's research can be disseminated. The Technical Paper Series publishes papers that are likely to be of specific interest to a relatively specialist audience, for example papers dealing with complex issues that assume high levels of prior knowledge, or those that make extensive use of sophisticated mathematical or statistical techniques.

The content and its format are entirely the responsibility of the author, and papers published in the Technical Paper series are not subject to peer-review or editorial control. Offers of further papers, and requests for information should be directed to Frances Sharp in the Publications Office, Centre for Health Economics, University of York.

July 2004.

Developing new approaches to measuring NHS outputs and productivity

First Interim Report

July 2004

Lead Investigators:

Diane Dawson[*], Hugh Gravelle[**], Paul Kind[*],

Mary O'Mahony[***], Andrew Street[*], Martin Weale[***]

With the assistance of:

Adriana Castelli[*], Rowena Jacobs[*], Sue Macran[*], Casey Quinn[*],

Philip Stevens[***], Lucy Stokes[***]

CHE Technical Paper Series No. 31

[*]Centre for Health Economics, University of York. **National Primary Care Research and Development Centre, Centre for Health Economics, University of York. ***National Institute for Economic and Social Research.

Contact points: methodology – hg8@york.ac.uk; mweale@niesr.ac.uk; outcomes: ads@york.ac.uk; inputs: m.omahony@niesr.ac.uk

Executive Summary

1. *Background*: In a recent review of the NHS Wanless (Wanless, 2002) highlighted the importance of obtaining good measures of NHS productivity for undertaking long-term forecasts of NHS resource requirements. As part of an international initiative to improve recording of public service output in national accounts, the Office of National Statistics has been seeking improvements to the way outputs of the health sector are measured. The lack of robust measures of output for all public services led to the appointment of the Atkinson Review with the remit of examining the future development of government output, productivity and associated price indices.

   As part of the initiative to improve measurement of NHS outputs and productivity, the Department of Health commissioned research from the Centre for Health Economics and the National Institute of Economic and Social Research to:

   i) Review and evaluate the existing literature on productivity measurement and identify conceptual and practical challenges to measuring productivity change in the NHS.
   ii) Investigate whether data are available to convert potentially relevant methodologies into measures of output and productivity change in the NHS.
   iii) Attempt empirical estimation of the most promising approaches to measuring productivity change subject to data availability.

   The research team will report on (i) by 30 July 2004, on (ii) by 30 November 2004 and on (iii) by 30 August 2005.

2. *Methodology*: It is important to distinguish between activities (operative procedures, diagnostic tests, outpatient visits), outputs (courses of treatment that may require a bundle of activities) and outcomes (the characteristics of output which are of value to individuals, such as health changes, waiting time, convenience, quality of facilities). In the measurement of private sector

productivity growth the focus is on outputs rather than the characteristics they produce because of the assumption that the market price of the output measures the consumers' marginal valuation of the bundle of characteristics from consuming the output. In measuring private sector productivity we also do not need to concern ourselves with counting activities because they are embodied in the outputs which are produced and sold.

In the NHS there are no final markets where patients buy outputs from producers. Since there are no prices to reveal patients' marginal valuations of NHS outputs, we have to find other means of estimating their value. We can do so in two ways: we can measure the outputs and attempt to estimate the marginal valuations attached to them or we can measure the outcomes produced by each unit of output and attempt to estimate marginal valuations of the outcomes.

Total Factor Productivity Growth (TFPG) is the difference between the growth rates of indices of the volume of outputs and inputs. TFPG will be underestimated if no allowance is made for the changing value to the consumer of outputs due to improvements in quality, such as greater health gains and reduced waiting times.

The Department of Health has traditionally measured productivity change by an index that weights activities by average unit cost (CWAI). This implies that costs reflect the value that society places on these activities at the margin. So, cochlear implant to treat deafness (at £23,889) is assumed to be fifteen times more valuable than a normal delivery in maternity care (at £1,598). The use of unit costs as weights reflecting the marginal social value of outputs rests on strong implicit assumptions unlikely to be valid for the NHS. Even under these conditions, marginal rather than average costs would be relevant and quality change is largely ignored. Despite these severe limitations, in the short run there may be no practical alternative to unit costs as weights in output indices for much of NHS output. In the long run we can hope to increase the proportion of NHS output measured using weights which more closely reflect the value of NHS activities.

3. *Output and outcome measurement:* The alternative to valuing outputs via market prices (impossible by definition in the NHS) or unit costs is to estimate the volume of different outcomes generated by the outputs and to value the outcomes. Health gain is the most obvious outcome from NHS activity. The aim is to measure the change in health state following an NHS intervention relative to what it would have been without the intervention. There are a number of challenges to measuring health gain. These include how to measure health status and how to attribute changes in health status to NHS activity. We outline these technical issues and present the results of an early investigation into different sources of information from which to estimate health gain. Using data from published studies that include estimates of changes in health state following a medical intervention, we provide illustrative examples of how information on outputs, unit costs, health gain, waiting time and in-hospital death rates can be combined in indices of outputs.

   A core issue is the extent to which estimates of productivity change are sensitive to whether activities are weighted by unit cost or by health gain, ideally adjusted for other outcomes such as patient satisfaction and waiting time. Our examples show that Coronary Bypass would receive a relatively high weight in a cost-based index, but less so in a health gain based index. In contrast, hip replacements and upper genital tract procedures receive greater weight in the health gain based index. A cost weighted index is less sensitive to inclusion of a mortality adjustment than to inclusion of a waiting time adjustment.

   To minimise the use of cost weights, a key research problem is identification of weights for each characteristic of NHS outputs that reflect social valuations. There are several possibilities for valuation of changes in health states measured as Quality Adjusted Life Years (QALYs). For non-health related outcomes such as reduced waiting times and patient satisfaction, estimates might be derived from discrete choice experiments. The feasibility of using these data will be examined in the next phase of this research.

4. *Inputs and growth accounting:* Estimating productivity growth requires accurate measures of changes in the volume and quality of inputs. Here there are less severe problems than for the output side, as inputs are generally purchased in the market. The project will follow the conventional approach of dividing inputs into three broad categories, labour, capital and intermediate inputs, with further division in each category to capture changes in the quality of inputs used. Productivity growth will be measured using the growth accounting method that subtracts cost share weighted growth in inputs from output growth. This method allows a decomposition of improvements in output to changes in the volume of inputs (e.g. number of NHS employees), the quality of inputs (e.g. greater use of highly skilled nurses) and residual productivity growth.

   The report outlines data sources and methods that can be used to measure aggregate input growth and its components. It highlights that there are plentiful data in some areas (labour force) but that finding reliable data for other components (drugs prices, medical equipment) will be challenging. Related to this is the need to consider whether market prices for inputs reflect social marginal valuations in the light of the fact that for some inputs the NHS is a monopsony buyer.

5. *Future developments:* The research plan for the remainder of the project is as follows. First, there will be an assessment of data availability to convert potentially relevant methodologies into measures of output and productivity change, with a report due on 30 November 2004. This will be followed by empirical estimation of the most promising approaches to measuring productivity change in the NHS, with a report due on 31 August 2005. At present it would appear that more data may be available to measure NHS outputs at the national level than for lower level organisations such as individual Trusts.

Table of contents

1 Introduction

Measuring the productivity of the NHS is not a simple task, as recognised in the recent Wanless review (Wanless, 2002) which highlighted the methodological and practical difficulties involved in measuring productivity in health care. These include:

- the lack of prices for the majority of outputs
- difficulties in measuring the quality of service
- difficulties in aggregating many types of activity
- difficulties in accounting for changes in the skill mix of the workforce

The series most commonly used to monitor productivity in the NHS is expenditure per unit of activity, the Cost Weighted Activity Index (CWAI). In recent years the series shows rising real expenditure per unit of cost weighted activity and has been used to suggest poor productivity performance. In contrast, one of the few studies that has attempted to measure labour productivity in the UK combined health and social services sector, which is dominated by health, shows growth rates in the 1995-2001 period higher than in the early 1990s and much higher than most other EU countries (O'Mahony and van Ark, 2003). This discrepancy shows that the method of measurement is central. Such concerns led to the setting up of the Atkinson Review on the measurement of public sector productivity and to the commissioning of this research project by the Department of Health. The Atkinson Review published an interim report on 19 July 2004 (Atkinson, 2004).

Productivity studies in general have difficulty in incorporating measures of quality change, in particular in service sectors including health. In this sector, improvements in the quality of the service produced are likely to have been an important source of productivity change. Development of measures of quality change in health care are important for management of many aspects of the NHS and should be central to the development of measures of productivity growth. Productivity cannot be measured without some means of first measuring and then valuing quality improvements.

It can be useful to measure productivity change at different levels:

- For the NHS as a whole for comparison with other parts of the public sector.

- By organisations within the NHS (such as Primary Care Trusts (PCTs), general practices, NHS hospital Trusts, or specialties within Trusts) for monitoring, comparing, and managing such organisations.
- Analysis might be conducted for groups of conditions subject to National Service Frameworks or individual diseases or conditions.

The prime aim of the research project is to investigate methods of measuring outputs and productivity in the NHS as a whole and to examine the feasibility of developing a workable measure of productivity change for the NHS. The research team will:
1. Review and evaluate the existing literature on productivity measurement to identify the conceptual and practical challenges in measuring productivity change in the NHS.
2. Investigate whether data are available to convert potentially relevant methodologies identified into measures of output and productivity change in the NHS.
3. Attempt empirical estimation of the most promising approaches to measuring productivity change identified in Stage one subject to data availability reviewed in Stage two.

The project team will produce three reports:

- Review of available methodologies (this report)
- Assessment of data availability to convert potentially relevant methodologies into measures of output and productivity change, 30 November 2004
- Empirical estimation of the most promising approaches to measuring productivity change in the NHS, 31 August 2005

A search of the published and grey literature was undertaken. Details of the databases and search strategy are given in Appendix A. Relevant material from the search has been incorporated into the text of this report. There may be on going international research on output measurement and productivity that has not yet appeared in the literature. The Atkinson Review team surveyed other national statistics organisations and we are seeking information from academic researchers in other countries.

In section 2 of this report we set out some of the methodological problems that arise in attempting to measure NHS outputs and productivity. Sections 3 and 4 discuss the

empirical issues in respect of outputs and inputs respectively. Our preliminary conclusions are in section 5.

2 Productivity measurement: methodological issues

2.1 Some definitions: activities, outputs, outcomes

It is useful to distinguish *activities* (operative procedures, diagnostic tests, outpatient visits, consultations…), *outputs* (courses of treatment which may require a bundle of activities), and *outcomes* (the characteristics of output which affect utility). The focus in health economics has been on the change in health produced by a course of treatment, typically measured in quality adjusted life years (QALYs). But other characteristics of treatment also affect utility: the length of time waited for treatment, the degree of uncertainty attached to the waiting time, distance and travel time to services, the interpersonal skills of GPs, the range of choice and quality of hospital food, the politeness of the practice receptionist, the degree to which patients feel involved in decisions about their treatment, etc. The distinction between outputs and outcomes is identical to that between goods and characteristics in consumption technology models (Deaton and Muellbauer, 1980, Ch. 10; Lancaster, 1971) where consumers value goods because of the bundle of utility yielding characteristics they produce. We can also think of the *quality* of the output as some function of the vector of outcomes it produces (see section 2.4).

In the measurement of private sector productivity growth the focus is on outputs rather than the characteristics they produce because of the assumption that the market price of the output measures the consumers' marginal valuation of the bundle of characteristics from consuming the output. In measuring private sector productivity we also do not need to concern ourselves with counting activities because they are embodied in the outputs which are produced and sold.

In the NHS there are no final markets where patients buy outputs from producers. This has two consequences for attempts to measure NHS productivity. First, because there are no prices to reveal patients' marginal valuations of NHS outputs, we have to

find other means of estimating their value. We can do so in two equivalent ways: we can measure the outputs and attempt to estimate the marginal valuations attached to them or we can measure the outcomes produced by each unit of output and attempt to estimate marginal valuations of the outcomes. The bundle of outcomes produced by a unit of output is likely to change over time in the NHS because of, among other things, changes in technology or treatment thresholds. In a private market the price of output would change to reflect this. But in the absence of market prices for NHS outputs it is likely to be easier to calculate the change in the marginal value of output by focusing on the change in the vector of outcomes. We discuss how the changing mix of outcomes (quality change) may be allowed for in section 2.4.

The second consequence of the lack of a final consumer market in NHS outputs is that some outputs are not counted at all or are poorly measured. Instead there may be data only on the activities and even these may be lacking in many areas of activity. We discuss the consequences of this in section 2.9.

2.2 Total factor productivity growth

If private markets are complete and competitive, prices reflect marginal utilities of the services to consumers and the marginal costs of provision. With some additional assumptions, the measurement and interpretation of productivity growth is then straightforward. Index the firms by i and denote the vector of j outputs from each firm at a time t as $\mathbf{y}_i(t)$. We index the goods by j. Let $\mathbf{z}_i(t)$ be the vector of n inputs (types of capital, labour and materials). $v_i(t)$ is a parameter which captures the state of technology at time t. The technology of firm i is described by the implicit production function

$$g_i\left(\mathbf{y}_i(t), \mathbf{z}_i(t), v_i(t)\right) = 0 \tag{1}$$

Assume that the technology exhibits constant returns to scale (CRS).

Differentiating (1) with respect to time gives

$$\sum_j \frac{\partial g_i}{\partial y_{ij}} \dot{y}_{ij} + \sum_n \frac{\partial g_i}{\partial z_{in}} \dot{z}_{in} + \frac{\partial g_i}{\partial v_i} \dot{v}_i = 0 \tag{2}$$

A profit maximising firm in a competitive market will choose $\mathbf{y}_i(t)$, $\mathbf{z}_i(t)$ to satisfy $p_j = -\theta_i \partial g_i / \partial y_{ij}$, $w_n = \theta_i \partial g_i / \partial z_{in}$, where $\theta_i = -p_1 / (\partial g_i / \partial y_{ij})$ is the Lagrange multiplier on the production constraint. We can rearrange (2) as

$$\sum_j \omega_{ij}^y \frac{\dot{y}_{ij}}{y_{ij}} - \sum_n \omega_{in}^z \frac{\dot{z}_{in}}{z_{in}} = \frac{\theta}{py} \frac{\partial g_i}{\partial v_i} \dot{v}_i = \omega_{i1}^y \left(\frac{\partial y_{i1}}{\partial v_i} \frac{v_i}{y_{i1}} \right) \frac{\dot{v}_i}{v_i} \tag{3}$$

where $\omega_{ij}^y = p_j y_{ij} \Big/ \sum_j p_j y_{ij}$, $\omega_{in}^z = w_n z_{in} \Big/ \sum_j p_j y_{ij}$

The left-hand side of (3) is the rate of change of a Divisia quantity index of outputs, minus the rate of change of a Divisia quantity index of inputs. Since total factor productivity (TFP) is the ratio of an index of outputs to an index of inputs, the left hand side is also a measure of total factor productivity growth (TFPG). If production takes place with constant returns to scale, then the total value of the product is expended on the costs of the inputs and we can replace the second term on the left hand side with the rate of change of an input index based on the cost shares $w_n z_{in} \Big/ \sum_n w_n z_{in}$

2.2.1 Significance of TFPG

The middle and last terms in (3) are equivalent expressions for the rate of technical progress. In the last term the rate of technical progress is given as the increase in one output (y_1), holding all other outputs and inputs constant, made possible by the change in technology. Thus TFPG also measures the rate of technological progress.

Technical progress increases welfare by relaxing the production constraint on the economy. Under certain assumptions total factor productivity growth can be given a direct welfare interpretation. Thus suppose that the economy is characterised by the implicit production function $g(\mathbf{y},\mathbf{z},v) = 0$ and resources are allocated to maximise current period welfare $U(\mathbf{y},\mathbf{z})$ where $\mathbf{y}$ and $\mathbf{z}$ are vectors of outputs and inputs. (A fuller treatment would consider the more complex intertemporal welfare problem but the basic result about the effect of a change in technology and its relationship with

TFPG would continue to hold.) The Lagrangean for the welfare problem is

(4) $$L = U(\mathbf{y}, \mathbf{z}) + \lambda g(\mathbf{y}, \mathbf{z}, v)$$

and from the envelope theorem

(5) $$dU / dv = dL / dv = \partial L / \partial v = \lambda g_v$$

Hence, if U is derivable from an individualistic, non-paternal welfare function, the fact that the allocation in an economy with a complete set of competitive markets maximises some such welfare function, means that TFPG is an increasing monotonic function of the change in welfare resulting from technological change.

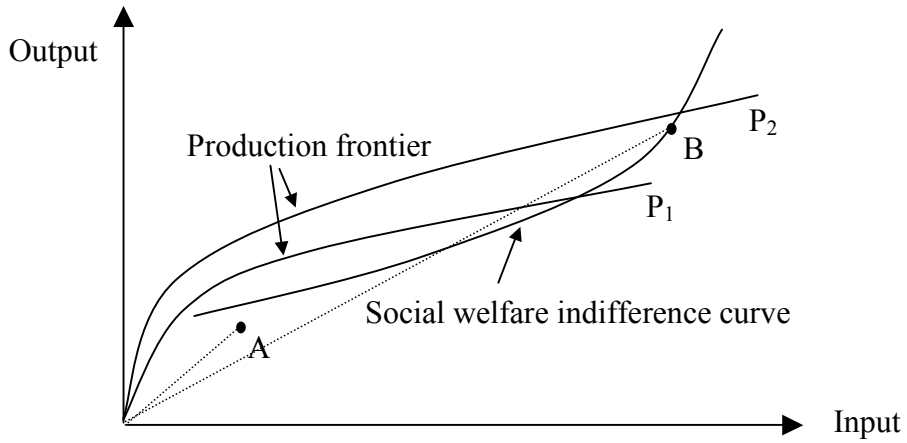
Typically much of the output growth is due to changes in input use, both changes in their volumes and in their quality. The growth accounting method employed to achieve this decomposition is outlined in section 4.1.

2.3 Application to the NHS

The application of these results to the NHS is problematic for two reasons. The first is that almost all NHS output is provided free of charge to consumers at point of use. Even in the few cases where the NHS does sell its output to the final consumer, as for pharmaceuticals prescribed by general practitioners (GPs) and dispensed to patients who are not exempt from payment, the price does not equal marginal cost. Hence some other means must be found of estimating the marginal social value of NHS outputs. We discuss the alternatives in section 3.5. There are also some difficulties in interpreting prices paid by the NHS for its inputs as measuring their marginal social opportunity cost (see section 2.12).

The second major difficulty is that it is by no means obvious that NHS production is socially optimal. It may be technically inefficient in the sense that it is possible to increase some type of output without increasing inputs or reducing some other output. It may also be producing the wrong mix of outputs.

Figure 2.1 Productivity, efficiency and welfare

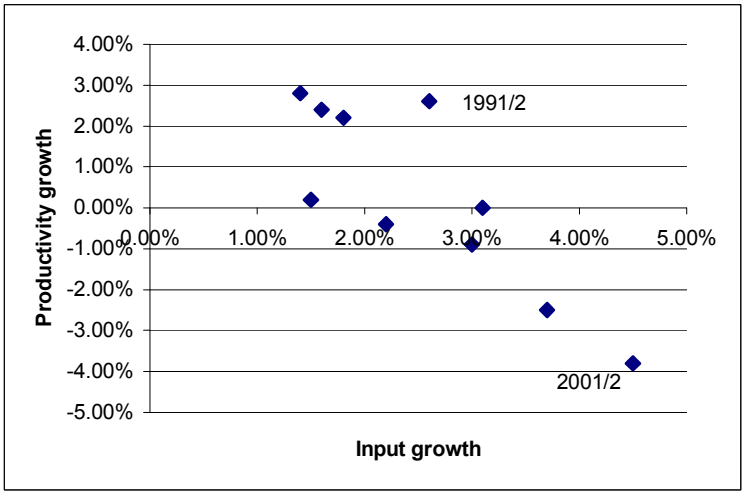


A at year 1 has higher productivity than B at year 2 but lower welfare and is less efficient (further away from its period production frontier)

Consider the simple single input, single output case in Figure 2.1. Point A in year 1 has higher productivity than point B in year 2 but welfare is lower at point A and, on any reasonable measure of technical efficiency, A has lower technical efficiency since it is further from its period production frontier. Technical progress has shifted the frontier upward from P_1 to P_2 but the productivity change does not even have the same sign as technical progress. The increase in welfare between period 1 and 2 is in part due to technical progress (B was not even feasible with the old technology) and to improvements in efficiency, perhaps because of changes in institutional structures and incentive mechanisms.

Note also that both technologies in this example have diminishing returns to scale so that increases in inputs along the frontier reduce productivity but that such a movement along the frontier can be welfare increasing. Figure 2.2 plots the rate of growth of NHS productivity, derived from the changes in the old version of the Cost Weighted Activity Index (see section 3.1) and an input index. The decline in the rate of NHS productivity growth may not indicate reduced efficiency or a decline in welfare but rather rapid increases in funding coupled with diminishing returns to scale.

Figure 2.2 HCHS productivity and input growth rates



Productivity growth: annual change in Cost Weighted Activity Index (old version). Input growth: annual change in base-weighted expenditure index. See section 3.1 for a description of contents of old CWAI.

These considerations suggest that there are problems in interpreting productivity growth as a welfare or efficiency measure. Nevertheless it can be a useful summary statistic to be used in conjunction with other data on the NHS. In section 4.1 below we outline the standard method to decompose output growth into contributions of inputs and productivity growth and argue that both are important in tracing improvements in the services provided by the NHS. A further justification for attempting to measure productivity is that it will stimulate improvements in NHS information collection and processing which may lead to improved decision making within the NHS.

2.4 Outputs, outcomes and TFPG

We drew a distinction in section 2.1 between outputs/goods and outcomes/characteristics and suggested that the distinction is crucial in the measurement of NHS outputs such as courses of treatment, and the outcomes, such as health gains and process utility, that these outputs generate. Consider first how we can construct a measure of TFPG in a market sector. Let the production function for a firm or sector which produces only one type (j) of output be

$$g_j\left(y_j, q_{j1}, \ldots, q_{jM}, \mathbf{z}_j, v_j\right) = 0$$

Here y_j is the volume or quantity of output j (the number of units produced) and q_{jm} is

the amount of outcome or characteristic *m* produced by consumption of one unit of output *j*. The vector $\mathbf{q}_j$ determines the quality of the product. At the equilibrium of a market economy the price paid for a unit of output *j* depends on the outcomes it produces: $p_j(\mathbf{q}_j)$, and is also a measure of quality. If the market for good *j* is competitive a profit maximising firm's choice of output, inputs, and outcomes will satisfy $p_j = -\theta \partial g_j / \partial y_j$, $y_j \partial p_j / \partial q_{ji} = -\theta \partial g_j / \partial q_{ji}$, and $w_n = \theta \partial g_j / \partial z_{jn}$. Totally differentiating the production function with respect to time gives

$$\frac{\partial g_j}{\partial y_j}\dot{y}_j + \sum_m \frac{\partial g_j}{\partial q_{jm}}\dot{q}_{jm} + \sum_n \frac{\partial g_j}{\partial z_{jn}}\dot{z}_{jn} + \frac{\partial g_j}{\partial v_j}\dot{v}_j = 0$$

and after using the profit maximising conditions, assuming constant returns to scale to substitute total cost for the value of output in the weights on the inputs, and rearranging we get

$$(6) \qquad \frac{\dot{y}_j}{y_j} + \sum_m \left(\frac{\partial p_j}{\partial q_{jm}} \frac{q_{jm}}{p_j} \right) \frac{\dot{q}_{jm}}{q_{jm}} - \sum_n \omega_n^z \frac{\dot{z}_{jn}}{z_{jn}} = \frac{\theta}{p_j y_j} \frac{\partial g_j}{\partial v_j} \dot{v}_j = \left(\frac{\partial y_j}{\partial v_j} \frac{v_j}{y_j} \right) \frac{\dot{v}_j}{v_j}$$

where $\omega_n^z = w_n z_{jn} \Big/ \sum_n w_n z_{jn}$.

Thus if we do not take account of the change in quality (the middle term in the left hand side of (6)) and merely calculate the difference between the rate of growth of the output and input indices we will not be measuring the rate of technical progress (the second and last terms). Equivalently, if we define TFPG as the difference between the rates of growth of the value of output and the cost of inputs, we will typically underestimate TFPG if we do not allow for the changing value of outputs because of improvements in quality. Consequently we need to take account of the change in the mix of outcomes (characteristics) embodied in each unit of output.

Denoting the marginal effect of outcome *m* on the price of output *j* as $\pi_{jm} \equiv \partial p_j / \partial q_{jm}$ we can write the rate of growth of the total value of output summed across all sectors ($Y = \sum_j p_j y_j = py$) as

(7) $$\frac{\dot{Y}}{Y} = \sum_j \frac{p_j y_j}{py}\left[\sum_m \frac{y_j \pi_{jm} q_{jm}}{p_j y_j}\frac{\dot{q}_{jm}}{q_{jm}} + \frac{\dot{y}_j}{y_j}\right] = \sum_j \omega_j^y \left[\sum_m \omega_m^j \frac{\dot{q}_{jm}}{q_{jm}} + \frac{\dot{y}_j}{y_j}\right]$$

where $\omega_m^j = y_j \pi_{jm} q_{jm} \Big/ \sum_m y_j \pi_{jm} q_{jm}$

In the competitive equilibrium these prices represent social values as well as costs of production. Thus, in principle the prices obtained in the competitive equilibrium enable us to calculate the rate of growth of the value of output and so derive the rate of technical progress via TFPG. In practice there are considerable difficulties even in market sectors in allowing for quality changes. In the NHS we have to estimate the marginal social value of the outputs (p_j) and effect of changes in the outcomes on these marginal social values (π_{jm}). If we are willing to assume that $p_j = \sum_m \pi_m q_{jm}$ we can replace the problem of estimating p_j with estimating the m marginal values π_m of the outcomes. The assumptions that the marginal social value of a unit of output j is a linear function of its characteristics and that the π_{jm} is independent of j characteristics are strong. The latter for example requires that an improvement in the quality of hospital food (say) per day in hospital has the same effect on the value of treatment for throat cancer as on the value of a hip replacement.

The discussion shows that a measure of TFPG which relates only to the volume of outputs and ignores their outcome or quality characteristics is incomplete. It also draws attention to the reality that health gains, typically measured in terms of QALY gains (see section 3.3) are only one of a number of outcomes produced by NHS outputs. It would be possible to measure all other outcomes in terms of QALY gains by converting them to QALY gain equivalents. But money provides a more obvious numeraire.

2.5 Changes in marginal social values over time

In section 2.4 we specified the value of NHS output as $Y = \sum_j p_j y_j$ $= \sum_j \sum_m \pi_m q_{jm} y_j$. In the rate of growth of the value of NHS output we assumed that the marginal social values of output (p_j) or of outcomes (π_m) were constant over time. If we allow for marginal values to vary over time then the rate of growth of the value

of NHS output is

$$(8) \qquad \frac{\dot{Y}}{Y} = \sum_j \frac{p_j y_j}{py}\left[\frac{\dot{p}_j}{p_j} + \frac{\dot{y}_j}{y_j}\right] = \sum_j \omega_j^y \left[\sum_m \omega_m^j \left(\frac{\dot{\pi}_m}{\pi_m} + \frac{\dot{q}_{jm}}{q_{jm}}\right) + \frac{\dot{y}_j}{y_j}\right]$$

Thus the rate of growth of the value of output depends both on changes in production conditions (the rates of growth of outcomes per unit of output and the rates of growth of outputs) but also on preferences (the rates of growth of the marginal social values of outcomes).

Under plausible assumptions the growth in the value of a QALY is determined by the rate of growth of income and the elasticity of marginal utility of income (Gravelle and Smith, 2001). But it is not affected by decisions within the NHS (except perhaps to a negligible extent because NHS decisions affect population health and thus the growth rate in income by improving worker productivity across the economy). Whilst changes in the value of a QALY and other outcomes may affect decisions about the allocation of resources within the public sector and the relative size of the public sector, it is not clear that they should be counted when we want to measure productivity growth i.e. the relationship between input growth and output growth. The terms involving the growth rate of the value of outcomes in (8) arise because we want to aggregate different types of outcome to calculate an overall index of outcomes. Thus there is a strong case for excluding them when measuring productivity, especially if the measure of productivity is intended to be used in part for monitoring the performance of the NHS.

Whilst we may want to exclude the growth in the marginal value of outcomes as contributing to TFPG we have to know whether and how the marginal values change over time in order to use the correct weights in calculating productivity growth

Note that

$$(9) \qquad \frac{\dot{p}_j}{p_j} = \sum_m \frac{\pi_m q_{jm}}{p_j}\left(\frac{\dot{\pi}_m}{\pi_m} + \frac{\dot{q}_{jm}}{q_{jm}}\right)$$

which again brings out the importance of the distinction between outcomes and

outputs. Even if we decide that the rate of growth of marginal social values should not be counted as part of productivity growth this does not mean that we should remove all of the rate of growth of marginal social value of *outputs* since part of $\dot{p}_j / p_j$ is due to changes in quality rather than to changing preferences.

2.6 Outcomes and attribution

Parts of the national income accounting literature note that health depends on factors in addition to health service outputs. For example health depends on income, education, age and other factors exogenous to NHS activity. Hence it is argued one cannot use outcomes to adjust outputs to take account of "quality" changes because changes in outcome may not be attributable to health service outputs. But what we want is the *marginal* effect of output *j* on health. If the health production function is additively separable in health service outputs and other factors the marginal effect of a health service output is well defined irrespective of the level of other variables affecting health.

It is more plausible that the health production function is not additively separable so that the marginal effect of y_j on health q depends on the confounding factors. This does not present a fundamental argument against the use of outcomes. The longstanding practice of using standardized mortality rates (SMRs) as a measure of population health suggests a way round the difficulty. Standardisation produces a measure of population health from which the effects of population structure (age and gender strata) have been removed so that one can make comparisons of mortality across periods or areas without the confounding effects of demographic structure. Under certain circumstances direct standardization can identify the true differences in mortality. The assumptions required are non trivial (age and gender specific mortality can be affected only proportionately by area or period (e.g. Yule, 1934) but direct standardisation is still useful. (The more common method of indirect standardisation which produces SMRs requires stronger assumptions.)

Thus, in the simple bilinear case, where health depends on a single NHS output y and say education x, the production function is

(10) $$h^t = a_0^t + a_1^t x^t + a_2^t y^t + a_3^t x^t y^t$$

and the marginal product of output y is $a_2^t + a_3^t x^t$. If $a_3^t = 0$ then we can set the health gain from treatment as $q^t = a_1^t$. But generally $a_3^t \neq 0$ and the growth in the marginal QALY effect of y is affected by changes in the confounding factor:

(11) $$\gamma^1 = \frac{a_2^1 + a_3^1 x^1}{a_2^0 + a_3^0 x^0}$$

To remove the effect of the confounding factor we can just choose an arbitrary level of the confounding factor x to replace x^t in (11). If we think that the changes in the coefficient a_3^t are not due to health service decisions then we should also standardize with respect to it as well:

(12) $$\gamma^1 = \frac{a_1^1 + \bar{a}_3 \bar{x}}{a_1^0 + \bar{a}_3 \bar{x}}$$

Obvious choices for $\bar{a}_3$ and $\bar{x}$ are their base period values or an average of the base period and current period values.

The health gains from treatment may increase simply because patients live longer. Consider the example of an increase in life expectancy that is not due to developments in the NHS but reflects rising living standards, changes in diet etc. As a result an NHS treatment, such as a hip replacement, may produce a greater outcome (QALY gain) because the recipient of a hip replacement is on average alive for longer to enjoy the reduced pain and increased mobility resulting from the procedure. Thus the marginal product (the QALY gain) of the treatment is greater for reasons arising outside the health service. The effect of longevity on the health gain from non-critical procedures is complicated.

The replacement hip is best seen as an investment from which the patient benefits for the rest of their life, at least if the hip does not wear out before death. Different patients have different lives and in aggregate the process can be seen as similar to one of radioactive decay - except that the decay rate is not uniform. Thus in aggregate we

can see death as the means by which the investment depreciates and an increase in life expectancy is a reduction of the rate of depreciation.

Considering the depreciation of a building provides an analogy. If the depreciation rate of buildings slows because of a change in the weather, no one would expect the volume measure of new buildings to be increased, although there might be derived effects leading to more buildings because they become, in effect cheaper. Unless the volume of gross output rises the output measure is unaffected. However, the reduced depreciation rate has the effect of raising net income relative to gross output. If we treated hip replacements as investments we would want to do the same. This approach is not being considered in the System of National Accounts (SNA) revision and is some way off. But the analogy indicates that, as far as possible, effects of changes to life expectancy which are quite independent of the procedures carried out should be kept out of the index.

There will be some cases where the gain in life expectancy may be partly due to improvements in the procedure and partly due to patients being better behaved- e.g. circulatory treatments go further if patients do not smoke. In terms of (10) the production function is not separable and judgement will be needed about how to unravel the impacts of factors exogenous to the NHS.

2.7 Outputs, outcomes and price indices

In the main example of a market based health care system – the US – the problems posed by quality change have led to attempts to refine price indices. The calculation of price indices for health care is very closely related to the problem of calculating health care output indices (it is the dual problem in a market economy). If health services are produced in the private sector, it is possible to calculate price indices which can be applied to the values of consumers' expenditure on health to get output volume indices.

One of the primary contributions of the US literature (see Appendix D for a brief account) has been to highlight the necessity of incorporating information on outcomes, even when health is privately provided and prices exist. Triplett (2001)

notes that that the interpretation of prices is less straightforward in health than in other service sectors such as car repair. The existence of asymmetric information means there may be a misalignment of the interests of patients and physicians, so that prices may not reflect consumers' willingness to pay. The problem is exacerbated by moral hazard arising from the fact that medical care is primarily financed by insurance rather than direct payments by the consumer. Hence there are arguments in favour of methods that focus on the direct measurement of the outcomes from medical interventions.

The approach is formalised in Berndt *et al.* (2000) in the context of deriving a cost of living index as an alternative to a more conventional services price index. The authors posit a stylised model based on a representative consumer choosing between consumption of goods and services (other than medical services) and health. Health in turn depends on medical interventions and a host of other factors such as lifestyles. Hence the output of the medical care industry should be seen as the *marginal* impact of health on utility, holding constant other factors affecting health, such as lifestyle. Berndt *et al.* (2000) derive a cost of living index which depends on health outcomes as well as other variables. Cutler *et al.* (2001) apply the approach to measure directly the cost of living index for the treatment of heart attacks.

Since there is no final consumer market in the NHS and hence no prices the question of how we could calculate price indices allowing for quality change is not directly relevant. But what is relevant is the conclusion that US researchers have drawn that in order to calculate meaningful price indices in order to deflate expenditure series, one needs measures of health sector outcomes, not just the volume of outputs.

2.8 Non standard outputs

Although the NHS is primarily a service for treating sick people not all of its outputs fit easily into this category and some are difficult to quantify and value.

2.8.1 Public health

The NHS undertakes a significant amount of public health activity which is directed at improving the health of the population as a whole, rather than improving the health of

specific individuals: e.g. healthy eating or anti-smoking campaigns. There are obvious problems in quantifying these outputs and their outcomes. In the 2004 Spending Review, the Public Service Agreement (PSA) between the Department of Health and the Treasury gave priority to increasing these public health outputs and reducing health inequalities (HM Treasury, 2004). However, within the time scale of this project it will not be possible to develop methods for measuring the NHS contribution to changes in public health outcomes.

2.8.2 Diagnostic activity

The NHS provides information to patients who are worried about their health. Relieving the anxiety of someone who presents with chest pains but only has indigestion is an outcome, just as improving the health of someone who presents with chest pains and has heart disease is an outcome. Thus we need to take account of the value of information produced by negative diagnostic tests for those who are not, *ex post*, ill and who do not go on to receive treatment.

2.8.3 Screening

Screening of asymptomatic patients can detect disease earlier and improve prognosis in true positives but because of imperfect sensitivity and specificity we also need to take account of the number of false negative, false positive and true negative cases and to value their effects on those screened. Since such screening services are also available in the private sector we will investigate their prices as a possible measure of the value of information to patients (see also section 2.10.1).

2.8.4 Training

The NHS invests in training its staff (for example via the NHS University). Half of NHS doctors are in training and divide their time between patient treatment and medical education. Many consultants devote part of their time to training junior doctors. In most cases training and patient treatment are joint products of NHS inputs and there are problems estimating the relevant marginal products. The Department of Health finances much of postgraduate medical education and now attempts to ringfence these costs from NHS funding. A comprehensive treatment of this issue is beyond the scope of the present research.

2.8.5 Research

The NHS funds a range of research. Information is a classic public good with major methodological and practical problems of valuation. In this study we do not attempt to measure the value of research outputs.

In considering non standard outputs, at most we can expect to consider the sensitivity of our baseline estimates to variations in assumptions regarding the costs to the NHS of undertaking these activities; see section 4.5.

2.9 Activities or outputs as the unit of analysis

2.9.1 Activities: institutional approach

NHS productivity measures have been based upon estimates of the number of particular types of activities (procedures, consultations etc) or the number of patients treated in various institutional settings. For instance, the previous version of CWAI incorporated measures of the activity undertaken in twelve different settings, such as acute, community and mental health hospitals, outpatient departments, general practices, dental practices and in community settings.

There are advantages to continuing within this framework. In instances where care for a patient with a particular condition is provided entirely within one setting, aggregation within the setting is equivalent to aggregation by patient pathway or disease group. It ensures compatibility with current NHS reporting systems and is likely to prove amenable to analysis at a disaggregated level. It can be a useful means for monitoring and managing lower level units within the NHS. Further, the approach would ensure consistency with other policy initiatives, most notably the Financial Flows reforms (Department of Health, 2002b).

The major disadvantage is that most patient cases pass through more than one institutional setting and their care requires several activities. Thus, for example, a patient who has a hip replacement will typically have been seen in general practice, in

an outpatient department, treated as an inpatient in hospital and received after care treatment from her general practitioner and from personal social services. Such care patterns can lead to double counting and make problematic the valuation of output of separate sectors contributing to joint production across sectors.

Current routine administrative data systems do not enable us to track the resource use associated with individual patients as they move along care pathways across settings. Even within institutional settings data may not be appropriately linked. For example, whilst there are very detailed data on types and quantities of different drugs dispensed to the patients of individual general practitioners, they are not linked to the individual patient or even to diagnostic group, so that we do not know who got what prescriptions or for what condition.

2.9.2 Outputs: patient-centred or disease-based approach

The bulk of NHS activities or services are delivered to individual patients with the aim of improving their health. But a disease or patient pathway approach has demanding data requirements. The approach is being investigated by US researchers (Berndt *et al.*, 2002; Berndt, Busch and Frank, 2001; Cutler and Huckman, 2003; Shapiro, Shapiro and Wilcox, 2001) (see Appendix D) and, in the UK, by the Office for National Statistics. It is probably the best way forward in the long run but is not fully implementable with the types of data available in the NHS in the short to medium term. We will be exploring whether it will be possible to use a small number of disease or patient groups as exemplars of the approach. Some of the data required, e.g. the number of patients by type of intervention, are readily available in the UK but other key sources of data that would be required to apply the US methodologies to this country may be missing or will require considerable search.

The relative advantages of the patient/disease group and institutional setting approaches depend on the degree of coverage, ease and timeliness of data collection; the dangers of double counting (for instance, where patients suffer multiple health problems); the ability to link to data on outcomes or prices; and the usefulness of the disaggregated measures (for instance, in changing behaviour). For the short to medium term the lack of properly linked routine data suggest that the measurement of

NHS output in fact will be based predominantly on the measurement of activities.

2.10 Valuation of outputs

2.10.1 Market prices

Under certain conditions the market prices for goods and services measure their marginal social value and hence can be aggregated for the construction of measures of the growth rate of output. One possible method of valuing NHS output might be to use prices from the private sector. Some NHS activities have close matches in the private sector. Some types of elective care are provided both in the private and public sectors. There are a few private sector general practitioners. Non-emergency ambulance transport is similar to a taxi service.

But there are strong arguments against the use of prices for private health care to value NHS output:
(a) Because of insurance much of the care purchased in the private sector is consumed beyond the point where its price equals its marginal value to the patient.
(b) Private health care outputs produce a different, and arguably more valuable, mix of outcomes (better quality hotel services, shorter waiting times) so that the price of private heath sector output may overstate willingness to pay for NHS output.
(c) The private sector does not produce the full range of NHS outputs, especially emergency care.

Some private sector prices may provide a useful comparator for valuations derived by other means for a subset of NHS activities but private sector prices do not seem suitable as output weights for the bulk of NHS output.

2.10.2 International prices

There is a precedent in cost benefit analysis for using world prices to value domestic output when domestic prices are absent or distorted. The rationale is that because trade could take place at world prices, they are legitimate measures of opportunity cost to the domestic economy. This option is not particularly useful in the valuation

of UK health care outputs. There is not a significant world market in health care. In the countries that do have published prices for health treatment, these tend to be administered prices subject to stringent domestic regulation or negotiation. It is highly unlikely that the relative prices observed in other countries will correspond to the relative value of NHS outputs.

2.10.3 Unit costs

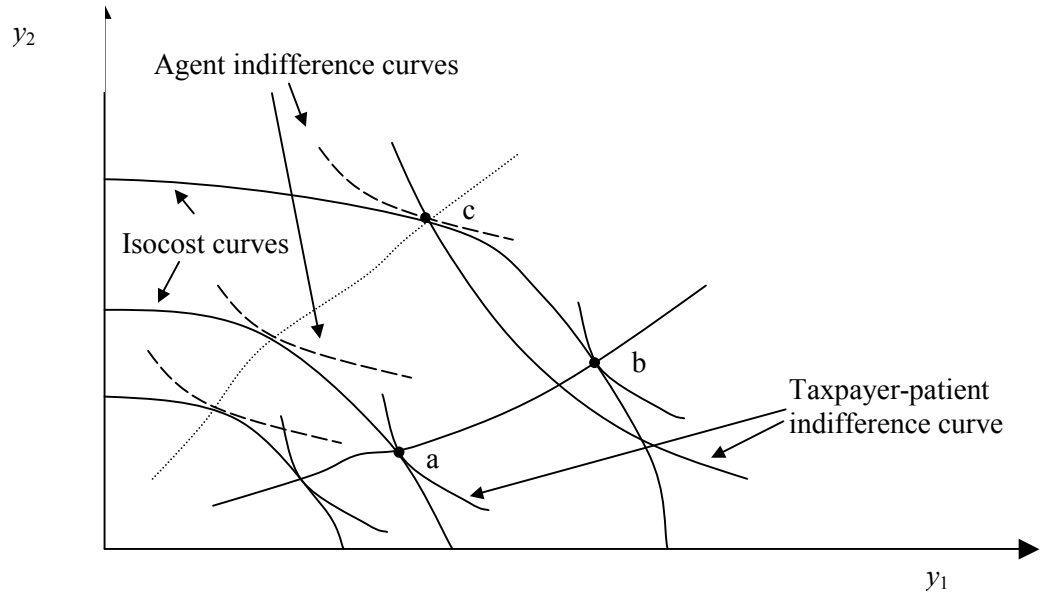
Current NHS practice, which follows the recommendation of Eurostat (2001), is to use production costs (such as the average costs as reported in the annually produced Schedule of Reference Costs) as weights in the calculation of output indices. This implies that costs reflect the value that society places upon these activities. So cochlear implant (at £23,889) is assumed to be 15 times more valuable than a normal delivery (at £1,598). The use of unit costs as weights reflecting the marginal social value of outputs has the support, albeit reluctant, of Hicks (1940) but it rests on strong implicit assumptions about decision making in the public sector. The use of unit costs raises a number of issues.

Imperfect agency. Suppose that there is a single taxpayer-patient with preferences over income x and two NHS outputs y_1, y_2 representable by the utility function

$$V = V^1(y_1, y_2) + V^2\left(x - c\left(y_1, y_2\right)\right) \tag{13}$$

where c is the cost of NHS output which is covered by a non-distorting tax. See Figure 2.3. The taxpayer-patient will choose an output mix at a satisfying $V_j^1 = V_y^2 c_j$, j = 1,2 so that the marginal value of output j at a is its marginal cost and $V_1^1 / V_2^1 = c_1 / c_2$.

Figure 2.3 Marginal costs as marginal values: imperfect agency



Suppose that decision making over the level and mix of NHS outputs is delegated to a political agent. If the agent has preferences over output and the costs falling on the taxpayer satisfy $G = G(V)$, $G'(V) > 0$, then she will also choose *a* and the marginal value of output is again its marginal cost. Now replace this perfect agent with an agent with preferences

$$G^b = V^1(y_1, y_2) + kV^2(x - C(y_1, y_2)),\ 0 < k < 1 \tag{14}$$

who thus respects the taxpayer's preferences over NHS output but places a lower weight on the cost consequences. Thus the taxpayer-patient and the agent's indifference curves in the output space in Figure 2.3 coincide. Output will be at *b* where $V_j^1 / V_y^2 = kc_j$ so that marginal cost of output *j* exceeds its marginal value to the taxpayer-patient. Notice however that at *b* the ratio of marginal costs equals the taxpayer-patient's marginal rate of substitution between outputs: $V_1^1 / V_2^1 = c_1 / c_2$. Hence an output index constructed using marginal cost weights would be proportional to one constructed using the taxpayer-patient's marginal valuations of the outputs and its rate of growth would, for small output changes, be unaffected by imperfect agency. Now replace this agent with one with preferences

$$G^c = G^{c1}(y_1, y_2) + kV^2(x - c(y_1, y_2)) \tag{15}$$

which give rise to dashed indifference curves in Figure 2.3. The allocation will now be at a point like *c* where the marginal cost of y_j does not equal the taxpayer-patient's marginal willingness to pay for it and $V_1^1 / V_2^1 \neq c_1 / c_2$. Hence marginal costs cannot be used to construct an output index with weights proportional to marginal willingness to pay and we cannot use the marginal costs to make inferences about whether a small change in the output mix makes the taxpayer-patient better or worse off.

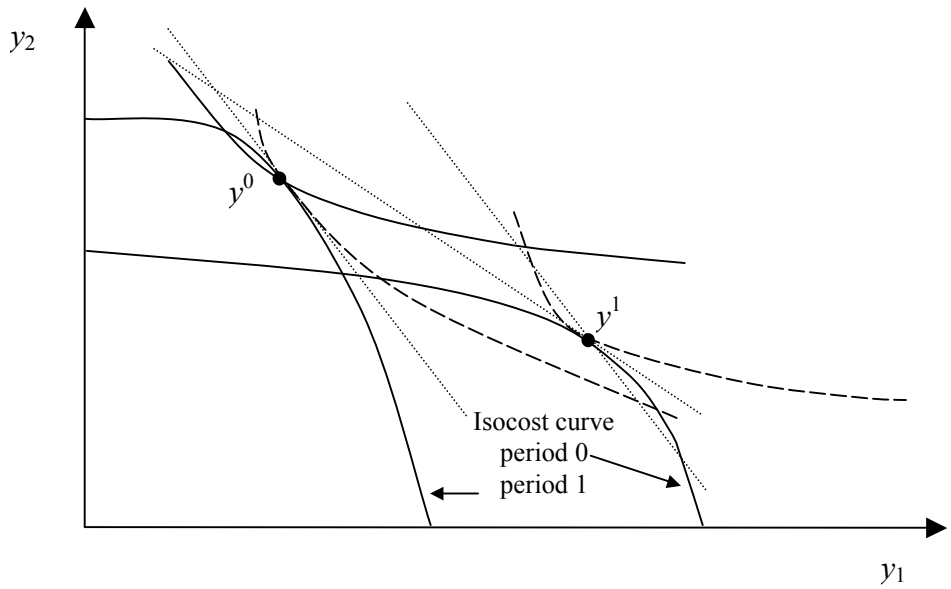
Figure 2.4 makes a similar point. The two isocost curves are for different periods and need not correspond to the same social cost. Social preferences are separable between outputs and cost, as in (13). When can we use information on marginal costs in the two periods to conclude that one output bundle is better than another ie that V^1 is increased or decreased by a move from y^0 to y^1? Suppose that the technology is convex so that the isocost curves are as shown in the figure. We have $\sum_j c_j^0 y_j^1 > \sum_j c_j^0 y_j^0$ where c_j^t is the marginal cost of output y_j in period *t*. Hence the base-weighted output index

$$\frac{\sum_j c_j^0 y_j^1}{\sum_j c_j^0 y_j^0} = \sum_j \left(\frac{y_j^1}{y_j^0} \right) \left(\frac{c_j^0 y_j^0}{c^0 y^0} \right) > 1 \qquad (16)$$

implies that y^1 was not feasible, with the expenditure and technology in period 0. But this is not sufficient to ensure that $V^1(y^1) > V^1(y^0)$. However if outputs are chosen to maximise an objective function for which preferences over output bundles are separable from cost and representable by V^1, then we can make more progress. We can then apply revealed preference arguments to use information on output indices with marginal cost weights to sometimes draw conclusions about the ranking of output bundles. (Though of course a full welfare judgement also requires information on the level of costs). Thus suppose that in Figure 2.4 the dashed indifference curves are contours of V^1 so that y^0 and y^1 maximise V^1 subject to a cost constraint. Then (16) still does not enable us to rank y^1 and y^0. But $\sum_j c_j^1 y_j^1 > \sum_j c_j^1 y_j^0$ so that the Paasche current marginal cost weighted output index exceeds 1 and we can infer that, because of the optimality of y^1 and y^0, y^1 must be on a higher contour of V^1 than y^0: $V^1(y^1) > V^1(y^0)$. But if the dashed indifference curves are those of an imperfect agent

choosing y we draw no conclusions from any output index weighted by marginal costs). For example if the solid indifference curve in Figure 2.4 is a contour of V^1 then $V^1(y^1) < V^1(y^0)$.

Figure 2.4 Imperfect agency and revealed preference



Deadweight loss from taxation. We assumed above that the NHS was financed by non-distorting taxation which is in contradiction to the first law of public finance: there is no such thing as a lump sum tax. The marginal deadweight loss from tax finance of public expenditure implies that the marginal social cost of public output is greater than its marginal cost of production. Hence, even if we believed that public expenditure levels were set such that the marginal valuation of public sector output was equal to its marginal social cost, the marginal cost of production would be less than the marginal social value of output. But the deadweight loss argument does not imply that we cannot use marginal cost weights to construct an NHS output index. Since NHS production is funded from general taxation, the marginal deadweight loss from taxation leads to the same proportionate difference between marginal production cost and marginal value for all NHS outputs. Hence the *ratio* of marginal costs of NHS outputs will be equal to the taxpayer-patient's marginal rate of substitution. Hence we can use marginal costs to construct an output index which is proportional to the output index which would be constructed using marginal social valuations of

outputs.

Inefficient rationing. Access to elective secondary care is rationed by waiting. There is a two stage process of producing care. First the supply capacity y^s is chosen. Then the capacity is rationed amongst patients. Let $B(y,w)$ ($B_w < 0$, $B_{yy} < 0$, $B_{yw} < 0$) be the net social benefit from elective care when the waiting time is w and y patients are treated. At any given waiting time, the rationing mechanism ensures that the marginal patient who joins the list generates a marginal benefit of zero ($B_y(y,w) = 0$) and that all infra marginal patients have positive net benefits. The demand for care is $y = D(w)$ ($D' = -B_{yw} / B_{yy} < 0$). In equilibrium $y^s = D(w)$ and the equilibrium waiting time is $w(y^s)$, $w' < 0$. The supply of care is chosen at the first stage to maximise $B(y,w) - c(y^s)$ subject to the second stage rationing mechanism. Hence the supply of care chosen satisfies

$$(17) \qquad d[B(D(w(y^s), w(y^s)) - c(y^s)] / dy^s = B_y D' w'(y^s) + B_w w'(y^s) - c'(y^s)$$

which implies, given the second stage rationing rule ($B_y = 0$), that the marginal cost of production satisfies

$$(18) \qquad c'(y^s) = B_w(y, w) w'(y^s)$$

Hence the marginal cost equals the marginal value of the reduced wait induced by an increase in supply. It does not equal the marginal value of treatment at the equilibrium wait which is zero given the rationing rule ($B_y = 0$). Rationing by waiting here imposes a deadweight loss on patients and does not generate any offsetting benefit to anyone else. But, despite the inefficient rationing mechanism, the marginal value of additional supply is measured by marginal cost and so marginal costs can be used as weights in an output index.

Quality change. Quality changes are pervasive in health care as result of technological change (new surgical techniques, new drugs), new methods of delivering services (NHS Direct, Walk in Centres), and new types of staff (Primary Care Mental Health Specialists, GPs with Special Interests). We can interpret such changes as leading to changes in the outcome vectors associated with an output.

Although such changes may have implications for the unit costs of outputs there is no reason why the value of the change in outcomes associated with an output should be captured by the change in the cost of production. Some technological developments are both cost reducing and quality improving. An increase in supply of elective care will, ceteris paribus, reduce waiting times, thereby improving the quality of care. But whether unit costs increase or decrease depends on the shape of the average cost curve. Similarly an increase in supply which leads to more treatment of patients with a given condition will be associated with a change in the average health gain unless patients are drawn entirely at random from the population with the condition.

In summary the above discussion suggests that the key assumptions required for marginal cost to be a measure of marginal social value of one output relative to another is that NHS agents respect the preferences of their taxpayer-patient principals in respect of the mix of NHS outputs and that quality changes are reflected in unit cost changes. The former is slightly less objectionable than the latter. Neither imperfect rationing nor the marginal deadweight loss are arguments against the use of marginal cost weights in an output index.

The costs used in the CWAI are average costs whereas the preceding discussion suggests that marginal costs should be used. It is extremely unlikely that average NHS costs equal marginal costs for all outputs. The requirement could be relaxed to the ratio of marginal cost to average cost being the same for all outputs. But this does not seem much more plausible given the presence of joint costs which are allocated to outputs using arbitrary accounting conventions in order to calculate unit costs. We will investigate whether it is possible to estimate cost functions in order to derive marginal cost weights (see also section 3.6). We suspect that at the moment there is no practical alternative for using unit costs as weights in output indices for much of NHS output. However the discussion above suggests that we should look for alternative weights and use them where possible.

2.11 Cost and QALY gains as weights

Suppose that we argue that the prime purpose of medical treatment is to cure people and other issues are secondary. Then the key outcome of interest is the QALY gain or

similar health improvement measure. Let q_j^t be the QALY gain outcome per unit of output *j* in period *t*. One method of allowing for "quality change" over time is to scale the level of output in period *t* by the growth factor $\gamma_j^1 = q_j^1 / q_j^0$ which is the proportionate change in the outcome per unit of output. We can calculate the outcome index for year 1 relative to year 0 as a weighted sum of growth rates of rescaled output:

$$(19) \qquad \sum_j \frac{q_j^1 y_j^1}{q_j^0 y_j^0} = \sum_j \frac{\gamma_j^1 y_j^1}{y_j^0} \frac{q_j^0 y_j^0}{\sum_{j'} q_{j'}^0 y_{j'}^0} = \sum_j \frac{\gamma_j^1 y_j^1}{y_j^0} \omega_j^{q0} = \sum_j \frac{\hat{y}_j^1}{y_j^0} \omega_j^{q0}$$

where ω_j^{q0} is the share of total QALYs generated by output *j* in year 0.

If we use cost weights we could allow for quality change by rescaling year 1 output:

$$(20) \qquad \sum_j \frac{y_j^1 \gamma_j^1}{y_j^0} \frac{c_j^0 y_j^0}{\sum_{j'} c_{j'}^0 y_{j'}^0} = \sum_j \frac{\hat{y}_j^1}{y_j^0} \omega_j^{c0}$$

where ω_j^{c0} is the share of total costs accounted for by output *j* in year 0 and c_j^0 is the marginal cost of output *j* in year 0.

Comparison of (19) and (20) shows that the two indices are equal if and only if

$$(21) \qquad c_j^0 / q_j^0 = \mu, \quad j = 1, ..., J$$

Thus the choice between QALY and cost weights does not affect the calculated rate of growth of output if (a) the NHS chooses its output mix as if it were maximizing total QALYs subject to a budget constraint which would imply that the marginal QALY gain from output *j* is proportional to the marginal cost of output *j*; (b) if q_j^t is the marginal QALY gain from output *j*; and (c) the cost weights c_j^0 are marginal costs.

Requirement (a) is perhaps the most demanding though with greater emphasis on cost effectiveness in the NHS following the establishment of NICE it may be more likely to be satisfied in the future. Most QALY gain data from interventions seems to be in the form of an average QALY gain and cost data are average and not marginal cost. If the average QALY gain is constant as output varies and if NHS accountants

allocate fixed costs in proportion to marginal cost then the two indices would again be identical.

2.12 Input prices

It is conventional to use the input prices paid by a sector as measuring the marginal social cost of its inputs. In competitive markets where buyers and sellers are price takers and suppliers of inputs bear the full costs of their decisions the convention is reasonable. However, in some input markets the NHS has considerable monopsony power. In others it bargains with powerful trade unions such as the BMA. In others, such as the pharmaceutical market, prices are determined by a complex regulatory mechanism.

In input markets where NHS purchases affect the market price because the supply curve is upward sloping, the marginal expenditure on the input will exceed the market price. Since NHS expenditure is financed from distorting taxation it should exercise its monopsony power in deciding on how much of the input to purchase and hence on the price to be paid. The amount bought will be less than if the NHS acted as a price taker. But suppliers will still be on their supply curve and, assuming the supply curve reflects all the costs of supply, the price paid by the NHS in a monopsonised market will equal the marginal social cost of the input.

Conversely, if sellers exert monopoly power and the NHS takes the price as given then the price paid will exceed the marginal social cost of supply. This is also likely to be true in markets, like that for doctors, where the NHS bargains with a powerful trade union. Depending on the specification of the bargaining model (for example whether we assume that bargaining is over the wage and employment, or just over the wage, with lower level decision makers in the NHS then acting as price takers in choosing employment levels) we get different predictions about the level of the wage and employment. However, except in the unlikely event that the NHS has all the bargaining power, so that we have a monopsonised market, the bargained wage will include some rent and so will exceed the marginal social cost of labour.

Around 30% of NHS Trusts expenditure is on pharmaceuticals. The prices that the

NHS pays for drugs are determined by a complex regulatory mechanism (the 1999 Pharmaceutical Price Regulation Scheme) which, roughly speaking, limits a firm's rate of return on drugs sold to the NHS. Although there is no direct regulation of prices, the rate of return constraint does, on average, constrain the price of the basket of drugs sold by each firm. In the absence of regulation the prices of drugs still in patent will exceed their marginal costs of production, so that the effect of regulation is to push prices closer, on average, to marginal production costs. In practice, prices are sufficiently above marginal production costs to enable firms to earn a rent which is intended to provide an incentive for research and development of new drugs. We suspect that it will not be possible to estimate marginal production costs for pharmaceuticals so that we will be forced to use the prices paid by the NHS. (See also section 4.2 where we discuss the implications of intermediate good prices for the attribution of productivity growth to sectors).

3 Output and outcome measurement

This section first considers previous attempts to measure the services produced by the NHS. It then considers how these methods might be improved, focusing on adjustments for quality change. These include taking account of health improvements and non-health aspects of quality change.

3.1 Cost weighted activity index and NHS productivity

The quantity indicator employed up to recently in the NHS was a Laspeyres chain linked cost weighted activity index (CWAI). Comparing period t with the previous period denoted t-1, this is given by Laspeyres chain-linked index with unit cost weights:

$$CWAI_t = \sum_j \omega^c_{jt-1} \frac{y_{jt}}{y_{jt-1}} = \frac{\sum_j c_{jt-1} y_{jt}}{\sum_j c_{jt-1} y_{jt-1}} \tag{22}$$

where ω^c_{jt-1} is the share of NHS expenditure on activity j, y_{jt} the volume of activity j

and c_{jt-1} the unit cost of activity *j* in period t-1.

One of the main problems with the original CWAI was that the categories of activity were very broad and reflected administrative convenience. The CWAI was developed using volume data on the 14 groups of activities in Table 3.1.

Table 3.1 Activities recorded in old Cost Weighted Activity Index

Inpatient and Day Cases – Total finished consultant episodes (all specialties) less well babies.
Outpatient, Accident and Emergency and Ward Attenders – Total outpatient attendances (all specialties), plus total A&E attendances, plus total ward attenders (all specialties)
Regular Day Patients – Total regular day attenders' attendances at NHS day care facilities.
Chiropody – Total face-to-face contact.
Family Planning – Total first contacts.
Immunisation and Surveillance – Total completed primary courses (diptheria, tetanus, polio, pertusosis, MMR and Haemophilus Influenzae), plus total reinforcing doses (diptheria, tetanus and polio) plus total number of BCG vaccinations.
Screening – Total number of cervical smears examined.
Health visiting – Total contacts by health visitors.
District Nursing – Total face-to-face contacts.
Community Psychiatric Nursing – Total face-to-face contacts.
Community Mental Handicap Nursing – Total face-to-face contacts.
Community Maternity – Total domicilary visits made by midwives (ante-natal and post-natal).
Dental – Total number of episodes of care.
Ambulance – Total number of patient journeys.

Finance data were obtained (from the programme budget) for each of the areas and were used to weight each of the sectors.

The CWAI did not take account of growth in areas where technological changes and treatment substitutions were important. Both these problems can be overcome to some extent by the more detailed activity index now available that combines unit costs and quantities for over one thousand categories of activity. These activities are grouped under a number of headings; activity growth rates and their corresponding expenditure shares are shown in Table 3.2. As can be seen from the Table, there is very large variation in growth across categories. As more data become available it will be

possible to move from a fixed base quantity index to a chain linked one, thus capturing substitution between treatments. However, to take account of quality change in this method it is necessary to collect information on the outcomes, as well as the output volumes, as in the recent price index studies carried in the US and reviewed in Appendix D.

Table 3.2 NHS activity growth 2001/2 to 2002/3

	Activity growth	Cost share
ALL	4.20	100
Elective Inpatient Including Day Cases (Including Primary Care)	4.56	13.07
Non Elective Inpatients (Including Primary Care)	1.93	18.99
Outpatient First Attendances (by Specialty)(Including Primary Care)	5.47	2.12
Outpatient Follow Ups (by specialty)(including primary care)	1.63	3.71
Total Outpatient HRG	4.87	4.93
Outpatient Maternity (incl primary care)	110.42	0.06
Mental Health (incl primary care)	-0.23	9.38
Accident & Emergency	4.48	2.39
Critical Care	-0.69	4.01
Radiology (including primary care)	-0.85	0.34
Audiological Services (including primary care)	1.10	0.07
Pathology (including primary care)	12.89	1.02
Chemotherapy	10.60	0.39
Renal Dialysis	-13.33	0.75
Bone Marrow Transplant	20.40	0.16
Spinal Injuries	11.36	0.11
Community Midwifery Services (including primary care)	11.99	0.46
Outpatient Community Services (including primary care)	-10.45	0.76
Rehabilitation (including primary care)	17.16	0.86
NHS Direct	10.87	0.32
Walk-In Centres	19.15	0.05
NHS Direct Online	95.84	0.00
Primary Care Consultations	10.22	12.92
Family Health Services Prescribing	7.90	16.74
General Dental Services	-0.61	4.61
General Ophthalmic Services	-1.49	0.82
Ambulance Emergency Journeys	2.85	0.94

DH calculated an index from the data underlying Table 3.2 and they also calculated

the new index back in time. The following table shows the results and the original CWAI. The striking feature of this Table is the different time profile of the old CWAI and the revised NHS outputs index. Old CWAI shows declining activity in the final few years but the new index shows increased activity. One obvious difference is that old CWAI is a Laspeyres chain linked index whereas the new outputs index is Paasche index using 2002/03 weights (from 2002/3 this will be a Fisher index). This may explain some of the difference. But recent ONS calculations based on similar data also show a discrepancy in the time profile between a CWAI using disaggregated data and the old CWAI.

Table 3.3 Output growth as measured by the old CWAI and the new NHS outputs index

Year	New outputs index	Old CWAI
1996/7	2.83%	1.70%
1997/8	1.83%	1.80%
1998/9	1.94%	2.10%
1999/0	3.36%	1.10%
2000/1	2.96%	0.60%
2001/2	4.41%	0.00%
2002/3	4.20%	n.a.

Note: Financial years

The very different time profile of the two indices shown in Table 3.3 suggests that the level of aggregation is an important consideration. How we slice up the available data can have important consequences for the final result. In fact aggregating some categories, e.g. treatments for a condition with similar outcomes, can be justified as partially taking account of quality change. This is the method used by many researchers in the US in developing price indices, reviewed in Appendix D. A similar approach applied to activity data is currently being developed by ONS. Hence a comprehensive study of the activities included in the new index, and how they relate to each other, will be an important contribution of the research project. Nevertheless, taking account of quality improvements requires a different approach and so is the subject of the remainder of this section.

3.2 Quality change: general remarks

The alternative to valuing outputs via market prices or unit costs is to estimate the volume of the different types of outcomes (q_{jm}) generated by the outputs and to value the outcomes (π_m). The Department of Health has looked for indicators of non-health characteristics of health care that are of importance to NHS patients. Surveys of the patient experience of NHS care have been developed using insights from past research, focus groups and interviews to identify aspects of the process of care of importance to patients. An improved index of changes in NHS outputs should incorporate changes in all the attributes of care valued by patients or the population. From existing evidence the main attributes to be valued are:

1. Improvement in health state
2. Waiting time
3. Choice of date for treatment and certainty of date of treatment
4. Nursing care
5. Time and communication with clinicians
6. Food
7. Physical environment (cleanliness, privacy, etc)

In this section we consider first the usefulness of existing data on changes in the outcomes and the extent to which these data may be employed in developing indices of NHS outputs. Valuation of outcomes is discussed in section 3.5.

3.3 Incorporating health improvements in productivity indices

Health gain is the most obvious outcome from the NHS. A treatment will alter the time stream of health related quality of life so that measurement of the health effect of treatment (the change in QALYs) requires measurement of changes in health over time and a discounting rule for adding up the health changes occurring at different points in time.

Florence Nightingale classified patients leaving her care as relieved, unrelieved or dead. Over one hundred and fifty years later the main NHS administrative data system for hospital patients is capable only of distinguishing between those discharged dead or alive. Death on discharge is a rare event, except for the very elderly (see Table 3.4) so that reliance on mortality data leaves us without any information on outcomes for the overwhelming majority of in patients who are alive

after their treatment.

Table 3.4 Dead on discharge from hospital 2002/3

Age group	Number of completed spells	Number discharged dead	Proportion discharged dead
0-1	905,373	5819	0.00643
2-5	291,437	199	0.00068
6-15	500,354	361	0.00072
16-25	1,019,578	935	0.00092
26-35	1,485,613	1743	0.00117
36-45	1,191,993	4123	0.00346
46-55	1,177,400	10034	0.00852
56-65	1,394,450	22206	0.01592
66-75	1,592,196	52533	0.03299
76-85	1,402,452	95603	0.06817
86+	535,219	71218	0.13306
All ages	11,461,096	264,161	0.02305

Source: Hospital Episode Statistics. Completed hospital spells including day cases. Discharged dead defined as *dismeth* = 4 or *dismeth* = 5. Completed spell as *dismeth* ≠ 8. Age defined by *startage*.

3.3.1 Measuring health related quality of life

The two key requirements for a measure of health status useful for measuring NHS productivity are a means of describing the health profile of an individual along a set of dimensions of health and a mapping from the set of dimensions to an overall health score. The basis of the present-day technology of health-related quality of life measurement can be traced back more than 30 years in the UK (Culyer, Lavers and Williams, 1971) and the United States (Fanshel and Bush, 1970).

There are two broad groups of health status measure. The first group is of condition-specific or targeted measures designed for use in a single therapeutic area. Even within a single diagnostic area such as depression, it is possible to find multiple measures. These measures often have their origins in clinical practice. The second, and much smaller group of measures, is generic and designed for use in a wide range of settings. Potentially such measures might be applied in all therapeutic settings. Practical design considerations mean that generic measures trade-off complexity and

comprehensiveness against reduced administrative burden and simplicity of coding. Since there are no absolute standards against which to judge the design of any measure, such compromises are reached on the basis of the experience of instrument developers and the technical performance of the measures they produce.

All health status measurement systems have two components. First, they require a means of describing health states. This can be done by an explicit formalised classification system (for example, the Karnofsky Performance Scale or Health Utilities Index) or, more often, by implicit combinations of responses to complex questionnaires (for example, Hospital Anxiety and Depression Scale or SF-36). Instruments differ in the dimensions of health they assess and the fineness of the measures employed along the dimensions. Second, all descriptive systems require some form of valuation or weighting system to reduce the volume of captured information to manageable proportions. Most weighting systems are characterised by arbitrary weights imposed by instrument developers. For example, responses to the 5-point Excellent, Very Good to Poor categorisation of self-rated health status in the original form of SF-36 is recoded to 100,75, …, 0. Following recommendations of the Washington Panel and the Technology Appraisal Guidelines published by NICE, it is widely held that the weights for health states (particularly in the context of cost-effectiveness analysis) should be based on those of the relevant general population. For the purposes of monitoring health outcomes at a system level, or at other than a patient level, this general principle seems well founded.

Weights for health states could be elicited from many different sources – health care professionals, patients or even health economists. It is sometimes argued that since patients have the relevant first-hand experience it is they who should be consulted about the issue of weights. However, such weights can be expected to vary in the course of the patient "journey" through ill-health; consequently, there is scope for seeing different values arising from patients with (say) newly diagnosed illness and those whose illness is of longer-standing. Similarly, patients who have wholly or partially recovered from an illness might hold different values to those whose response to treatment has been less positive.

The requirement for a standard metric to measure health outcomes at system level

rules out the consideration of all condition-specific measures based as they are on different descriptive systems each with an independent valuation or scoring system. Profile measures (such as the Nottingham Health Profile and SF-36) report generic health status in terms of separate dimensions of health but require a set of weights for the different dimensions to produce a single health index. The set of generic measures which produce a single index is small and includes Health Utilities Index, EQ-5D, 15-D and AQLQ. Of these measures, only EQ-5D is calibrated in terms of social preference weights of a UK population.

3.3.2 Measuring changes in health status

Given a suitable instrument to measure health status we must consider the baseline against which the post-treatment health status is compared in order to compute the outcome (the change in health status).

The two obvious baselines are:

- Pre-treatment health status. Let h_{jt-1} be the pre-treatment and h_{jt} the post-treatment health status so that the measured outcome is $\Delta h'_j = h_{jt} - h_{jt-1}$
- The health status that the patient would have experienced had there been no intervention. Let h_j^0 be the health status if the individual had not been treated. then the outcome is measured as $\Delta h''_j = h_{jt} - h_j^0$

The choice of baseline will yield different estimates of the health effect of interventions. In turn, this will lead to different estimates of productivity change over time if there is a change in the mix of activities from one period to the next.

To illustrate, consider two interventions *a* and *b*, with the same cost, suitable respectively for two individuals *A* and *B*, who suffer different conditions but who, prior to the intervention, have similarly poor health status, $h_{jt-1} = 0.5$, where *j=a,b*.

For individual *A*, intervention *a* yields no change in health status at time *t* relative to *t*-1, hence $\Delta h'_a = 0$, as shown in Figure 3.1. For individual B intervention *b* delivers an improvement in health status such that $h_{bt} = 0.7$, hence $\Delta h'_b = 0.2$. On this basis a

shift in the mix of activities from *a* to *b* over time would imply an increase in productivity if the activities had the same cost.

The true effect of the intervention is the change in health status with and without the intervention. This is not measured by the change in health status before and after treatment unless the individual's health without treatment would not have deteriorated (or improved). Suppose that the natural, untreated, course of disease for conditions *a* and *b* differs. Untreated, the health status on individual *B* would not change from one period to the next, with $h_b^0 = 0.5$ shown by the dashed line in Figure 3.2. The net treatment effect, therefore, amounts to $\Delta h_b'' = 0.2$. Individual *A*, however, suffers a debilitating disease, likely to result in a major deterioration in health status by time *t*, equivalent to $h_a^0 = 0.2$ if untreated. The role of the intervention in this case is not to improve health status, but to stabilise the condition. If the value of interventions is assessed on the basis of this net treatment effect, intervention *a* would be more highly valued with $\Delta h_a'' = 0.3$.

The obvious difficulty with estimating the effect of treatment, compared with no treatment, is that at date *t* every individual is either treated (so we don't observe the no treatment outcome) or not treated. What matters for estimates of productivity growth however is whether estimates of the *change* in the QALY gain from treatment derived from before and after comparisons is a biased estimate of the true change in the QALY gain. If it is possible to control for patient mix then changes in difference between pre- and post-treatment health status may be a useful proxy for the true change in the health gain.

The estimates derived by looking at the health status of individuals pre- and post-treatment has the merit of simplicity. We describe below a data set in which pre- and post-intervention measures of health status are available for individual patients treated in routine NHS practice.

Figure 3.1 Change in health status: before and after intervention

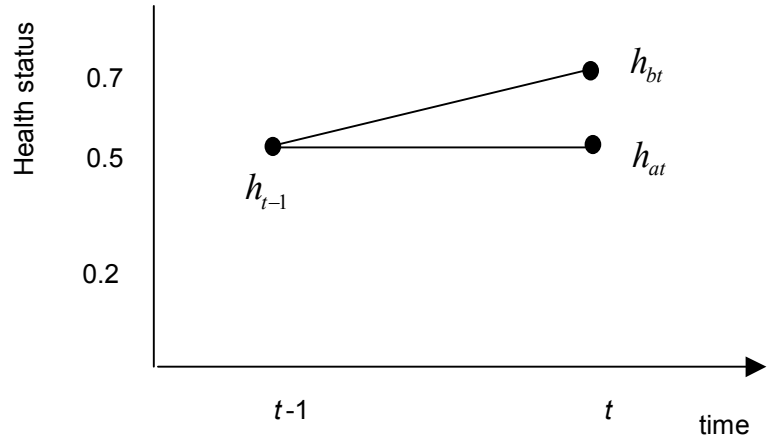
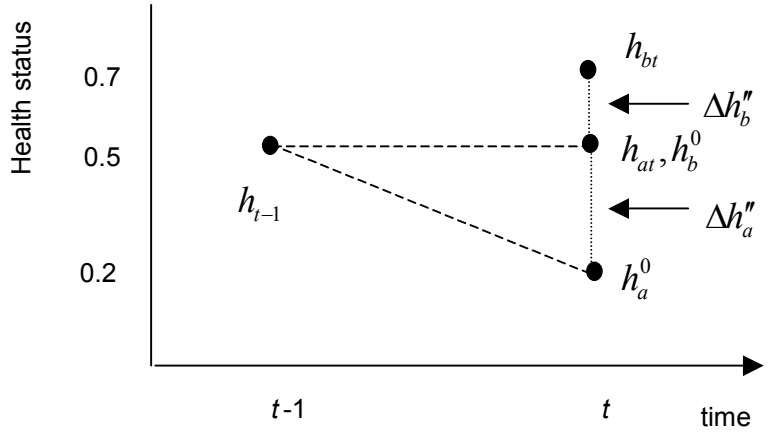


Figure 3.2 Change in health status: with and without intervention



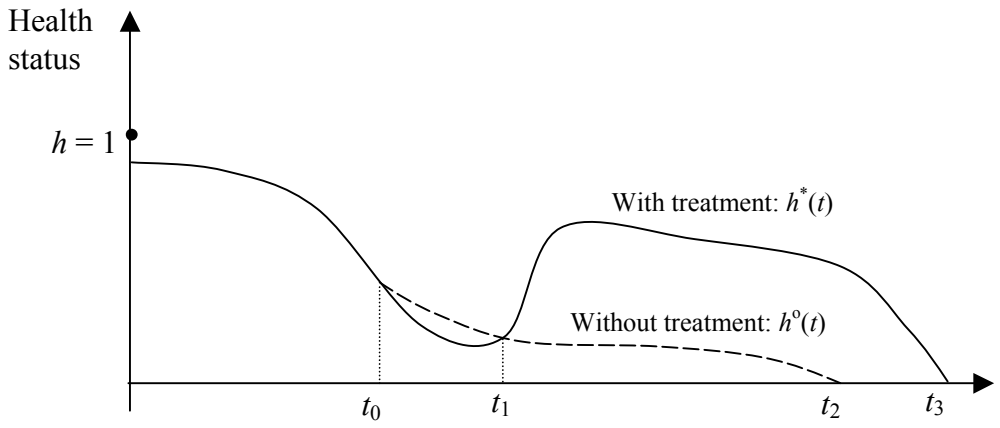
3.3.3 QALY gain calculation

A treatment will typically alter the future time path of health status so that it is necessary in principle to measure health status continuously in order to measure the full effect of treatment and hence to calculate the health outcome. In the example in Figure 3.3 treatment at time t_0 initially reduces health (for example the patient may be temporarily incapacitated even by successful surgery), but then increases health and lengthens life from t_2 to t_3. The Quality Adjusted Life Year gain from treatment is the integral of the difference between the time paths of health status with $h^*(t)$ and without $h^o(t)$ treatment.

Typically however we do not have estimates of the time paths of $h^*(t)$ and $h^o(t)$ but only of snapshot estimates at particular points in time. The timing of the snapshots is clearly crucial in Figure 3.3. In other cases it may be less of a problem because the difference between $h^*(t)$ and $h^o(t)$ is constant after a short time has elapsed after the treatment.

In some cases the effects of changes in mortality on the QALYs resulting from any particular procedure can be allowed for in a simple way by assuming that the QALY gain increases in line with the life expectancy of the patients treated. Although this is not ideal, it is better than ignoring the issue.

Figure 3.3 QALY gain from treatment



3.3.4 Source of information

3.3.4.1 Expert groups

It may be necessary to use "expert opinion" to generate data on health outcomes. This approach was taken by the World Health Organisation in their assessment of health system performance (World Health Organization, 2000), in recent US studies of productivity (Berndt *et al.*, 2002) and in early English studies of output measurement (Gudex *et al.*, 1990; Williams, 1985). Using the method to elicit outcome data could prove a formidable task but we have experience of implementing a viable approach among GPs that could be applied in this project (Kind *et al.*, 1993).

It may also be possible to piggy-back on other exercises. For example, The NHS Information Authority (NHSIA) has convened a set of Expert Working Groups (EWGs) in order to refine Healthcare Resource Groups and other NHS classification systems. EWGs would be ideally placed to undertake the role of estimating the health outcomes associated with HRGs. Supported by staff from the NHSIA, each EWG consists of clinicians with a demonstrable interest in improving NHS classification of activity and a fundamental understanding of the purposes and design of HRGs. Those involved may be expected to be reasonably well-disposed to the consideration of the health outcomes associated with each HRG.

Nevertheless, this would not be a trivial undertaking. The EWGs are working to a tight timescale and including the assessment of health outcomes would represent something of a deviation from their existing terms of reference, so it would not be a small matter for this to be included as a "bolt-on" to their existing commitments. Even so, it may be possible to have preliminary discussions with one or two of the EWGs in order to scope out the issues involved.

3.3.4.2 Clinical trials

The use of published studies to estimate changes in health status is not without precedent (Berndt *et al.*, 2001) and may be combined with or validated by expert opinion (Berndt *et al.*, 2002). Nevertheless, there are some drawbacks to using data from published studies. First, the study population may not be representative of the patients who receive the treatment in routine practice, given the exclusion criteria for

many studies. Second, the description of treatments in such studies may not map well to the classification of activities in the index. This is likely to be the case for the example below, where fairly precise treatment definitions have been mapped to the considerably more aggregated HRG descriptions. Third, the follow-up time in studies is variable, and hence the full treatment effect may not be captured by the estimate.

3.3.4.3 Observational data

An alternative to evaluation studies is systematic routine collection of QALY change data from patients before and after treatment. That it is possible routinely to collect health status data on patients has been demonstrated in two contexts. BUPA has been administering SF36 routinely to patients before and three months after treatment for a number of years (Vallance-Owen and Cubbin, 2002). For patients having eye operations the shorter VF14 instrument has been used. Around 100,000 patients have been surveyed. It may be possible to construct QALY gain scores from the pre- and post-treatment SF36 results but BUPA patients are not typical of NHS patients.

The second demonstration of the feasibility of collecting routine health status data is the Health Outcomes Data Repository (HODaR) (www.cardiff-research-consortium.co.uk/hodar). HODaR was set up under the aegis of the Cardiff Research Consortium, a joint venture of Cardiff University and University of Wales College of Medicine. Its principal objectives include the provision of health outcomes data to a number of potential interest/user groups, including the pharmaceutical industry. HODaR is a prospective survey of subjects treated as in-patients or out-patients at the third largest Hospital Trust (including University Hospital of Wales) in the UK and has been in operation since June 2002.

The HODaR database includes anonymous data on patients collected from three hospitals in the geographical region of Cardiff and the Vale of Glamorgan, South Wales. Survey data cover socio-demographics, resource use, productivity costs and health related quality of life. Electronic hospital data – such as the hospital episode statistics - are available for all responders, and these data linked with the survey returns. HRG version 3.5 codes based on patient diagnostic and operation data are generated by the participating hospitals and subsequently based to HODaR. Each patient is represented by a unique identifier; each in-patient admission or out-patient

clinic visit is associated with a unique event code. Health-related quality of life data are captured using the RAND-MOS SF-36 and EQ-5D. We report here some of the data on EQ-5D scores. In-patients are sent a postal questionnaire 6 weeks post-discharge. Out-patients are surveyed at the time of their clinic appointment. At the time of writing, a total of 32,268 events are logged within HODaR.

Table 3.5 Summary EQ5D data from HODaR

HRG v3.5	Description	N	EQ5D-PRE	EQ5D-POST	Change in EQ5D	Std Dev
L21	Bladder Minor Endoscopic Procedure w/o cc	118	0.705	0.677	-0.028	0.193
B13	Phakoemulsification Cataract Extraction and Insertion of Lens	84	0.654	0.610	-0.043	0.199
F06	Oesophagus - Diagnostic Procedures	84	0.647	0.626	-0.022	0.208
F35	Large Intestine - Endoscopic or Intermediate Procedures	75	0.647	0.636	-0.012	0.281
A07	Intermediate Pain Procedures	45	0.216	0.223	0.007	0.232
J37	Minor Skin Procedures - Category 1 w/o cc	38	0.658	0.610	-0.048	0.173
E14	Cardiac Catheterisation without Complications	31	0.636	0.596	-0.040	0.176
S22	Planned Procedures Not Carried Out	25	0.637	0.641	0.004	0.294
S27	Malignant Disorder of the Lymphatic/ Haematological Systems with los <2 days	21	0.540	0.568	0.027	0.256
E22	Coronary Atherosclerosis >69 or w cc	20	0.577	0.476	-0.101	0.244
E04	Coronary Bypass	19	0.481	0.603	0.122	0.199
L19	Bladder Intermediate Endoscopic Procedure w/o cc	19	0.680	0.755	0.076	0.229
E29	Arrhythmia or Conduction Disorders >69 or w cc	17	0.700	0.702	0.002	0.086
E35	Chest Pain >69 or w cc	17	0.440	0.385	-0.055	0.295
H04	Primary Knee Replacement	17	0.422	0.516	0.093	0.232
H10	Arthroscopies	17	0.546	0.587	0.042	0.180
E15	Percutaneous Transluminal Coronary Angioplasty (PTCA)	16	0.654	0.689	0.035	0.311
M07	Upper Genital Tract Major Procedures	16	0.648	0.664	0.016	0.217

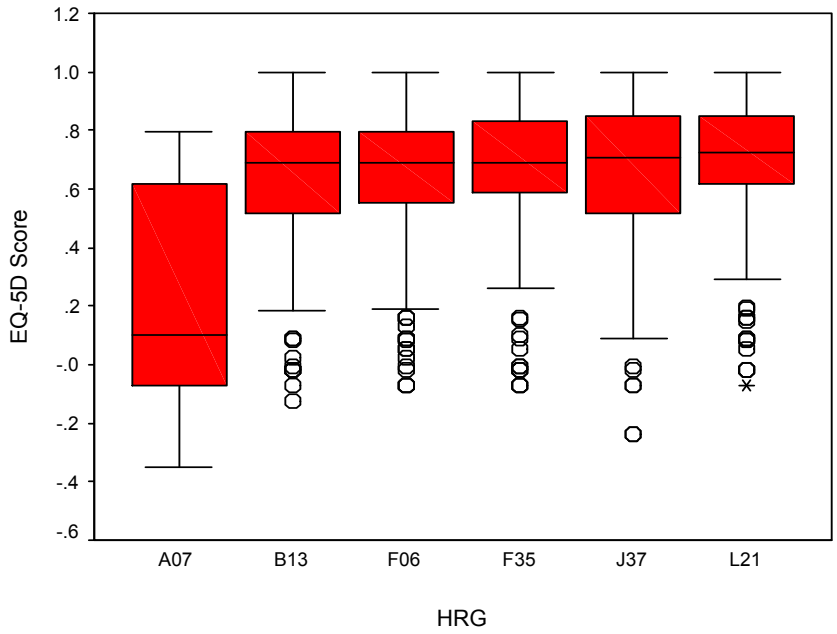
The major drawback with the HODaR data is that health status is measured only once during an episode of care, typically post care. Thus it does not provide snapshot estimates of the effect of care. But some patients in the database have been treated more than once and so for 2,360 patients there are two measures of health status. Such patients are unlikely to be typical of patients receiving care but we have undertaken some exploratory analysis of data for this sub-group.

The data were analysed according to the HRG to which the patient is coded at the second event. The coding classification for an individual patient may change during

the course of treatment, particularly for patients suffering multiple conditions. Summary EQ5D data for those HRGs containing more than 15 observations are presented in Table 3.5.

HRGs are designed to be resource homogenous. They may not be homogenous in relation to patients' health status, however. The variation in post-intervention EQ5D scores for patients classified to each HRG is illustrated for the six most populated HRGs, in the box-and-whisker plot below (Figure 3.4). These plots show the median (the horizontal dark line), the interquartile range (the shaded box), the 95% confidence interval (the "whiskers"), outliers (the circles) and extreme values (the stars). For HRG A07 (intermediate pain procedures involving a length of stay under two days), variation in EQ5D scores is considerable, suggesting heterogeneity in the health status of patients classified to this HRG. But it may be related also to there being only 45 patients for whom EQ5D scores are available for the HRG. The variations are less pronounced for the other five HRGs, all but one of which are based on larger samples.

Figure 3.4 Variation in post-intervention EQ5D scores for six HRGs



The greater the variation in EQ5D scores for patients classified to a particular intervention category (in these cases, defined by HRG), the less confidence there can

be in using a single estimate of the likely health outcome. In the presence of large variations, productivity indices will be sensitive to the characteristics of the sample from which the health outcome data are drawn. Variation is likely to increase the more aggregated the system of classifying interventions. HRGs, for instance, group together a large number of different procedures or diagnoses – for instance, Coronary Bypass (HRG E04) covers 52 procedures. It would not be surprising were the health outcomes to be quite different for patients undergoing different procedures within the more general classification.

Within an intervention category, two potential sources of variation might be investigated. First, some of the variation may be due to observable characteristics of the patients, over and above the effect of the intervention. If outcomes appear to be influenced by (say) the age or gender of the patient, estimates of the intervention effect can be conditioned upon these factors. The second source of variation might be more difficult to identify and correct for. Technical and allocative inefficiency may well pervade the NHS (variations in clinical practice might be cited in evidence). If so, there may be an institutional effect that influences the health outcome data from which the sample is drawn. This may caution against relying on sample data from a single institution. These issues both require further consideration.

The pre- and post-intervention HODaR data were used to calculate the change in health status. Figure 3.5 plots the mean change in EQ5D scores for all HRGs, ordered first according to the number of observations, and then by the size of the change, with the most populated HRGs appearing at the left of the scale. Two messages emerge. First, the scale is centred - more or less - around zero. This implies that the improvements in health status experienced by patients in some HRGs are cancelled out by deteriorations in health status experienced by others. Second, larger changes in the mean are apparent as the sample size decreases – the HRGs at the right end of the scale contain only single observations. After excluding HRGs that contain fewer than ten observations, the variation in the mean change in EQ5D score is lessened considerably, as shown in Figure 3.6.

Figure 3.5 Average change in EQ5D score by HRG

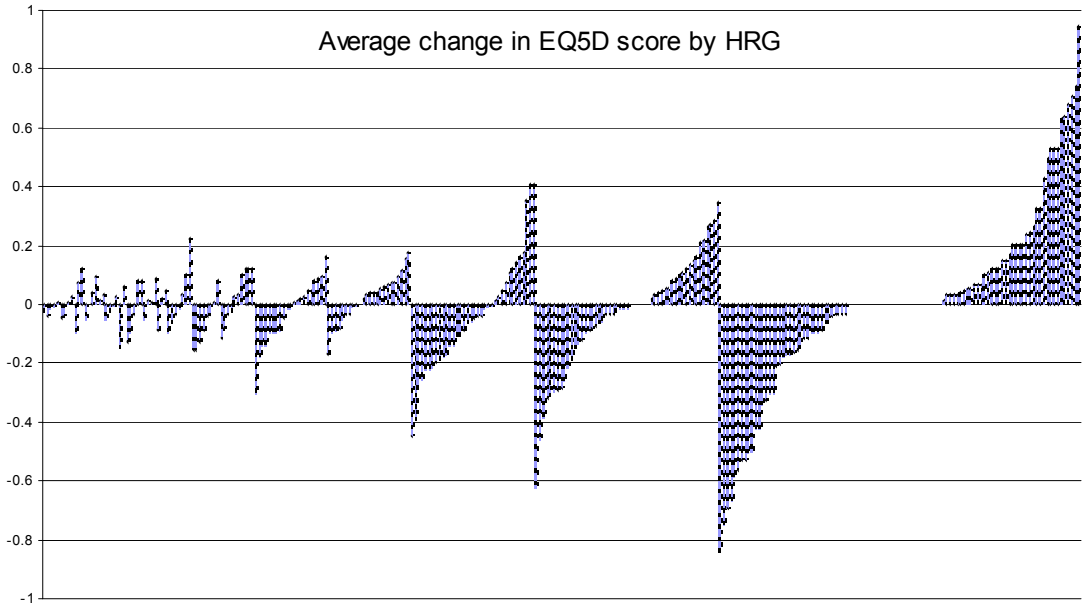
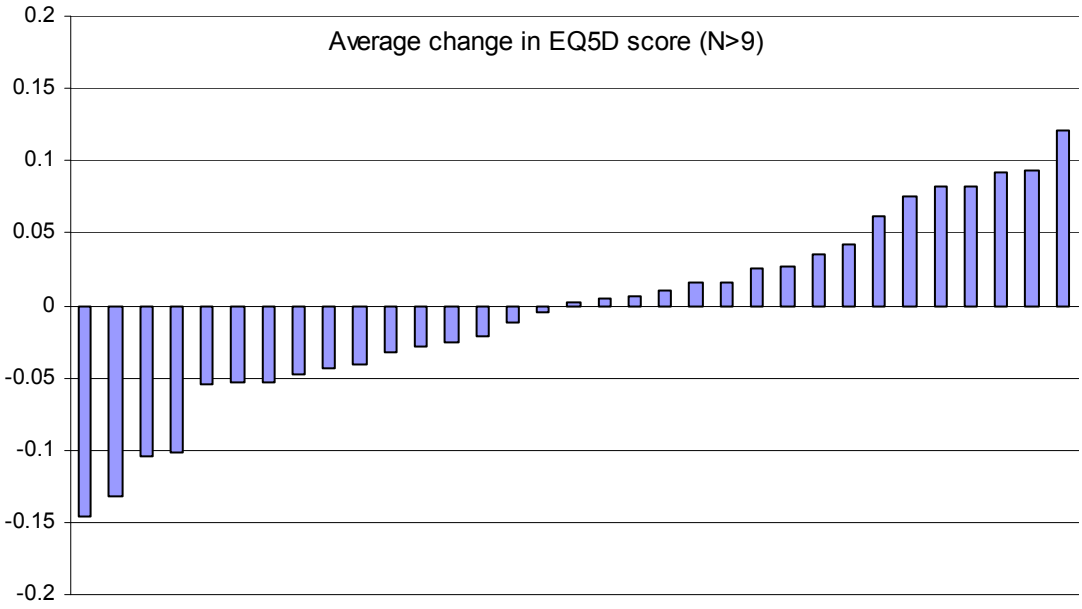


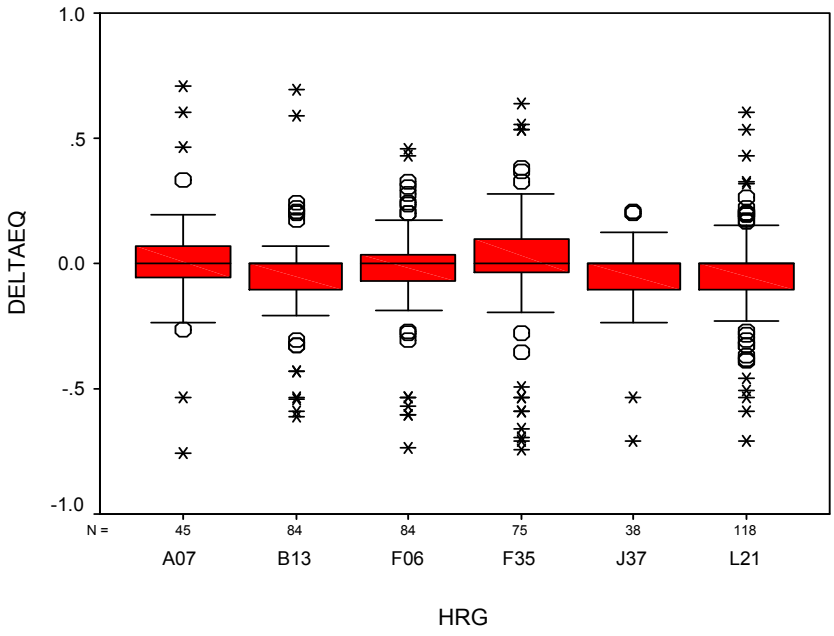
Figure 3.6 Average change in EQ5D score for HRGs with more than ten observations



Finally, returning to the six HRGs with the most observations, Figure 3.7 presents box-and-whisker plots to show the within-HRG variation in the change in EQ5D scores. For these the median change in health status is zero. This raises the questions of, firstly, whether it is appropriate to measure change in health status in this way,

rather than in relation to expected health status in the absence of intervention and, secondly, whether EQ5D is sufficiently sensitive to detect changes in health status.

Figure 3.7 Variation in the change in EQ5D scores for six HRGs



3.3.5 Illustrative example

In this section we give some examples of how information on outputs, unit costs, QALY gains, waiting times and in-hospital death rates can be combined in indices of outputs. The formulations and calculations are purely illustrative. We will be investigating other index number formulations, with different implicit weights on the components, and with a much wider set of outputs. (Recently the DH has been exploring the possibility of incorporating into its output and productivity indices a range of information on waiting times, mortality rates and other outcome indicators, though not QALY gains.)

3.3.5.1 Method

Rather than using the HODaR data, we identified a small number of studies which report *changes* in health status as measured by EQ-5D (Bosch, van der Graaf and

Hunink, 1999; Brilstra *et al.*, 2004; Eefting *et al.*, 2003; Garry *et al.*, 2004; Lloyd, in press; Manca *et al.*, 2003; Nathoe *et al.*, 2003; Ostendorf *et al.*, 2003; Prinssen, Buskens and Blankensteijn, 2004; Sculpher *et al.*, 2004; Tangelder *et al.*, 1999) and identified the version 3.1 HRG that would best describe the patients included in the study. (Version 3.1 HRGs rather than version 3.5, as coded by HODaR, are used because the former classifications are the basis of reference cost returns). This generated a set of nine HRGs, listed in Table 3.6, with details of the number of finished consultant episodes in 2002/03 and the change in QALYs resulting from treatment. The following data were collected:

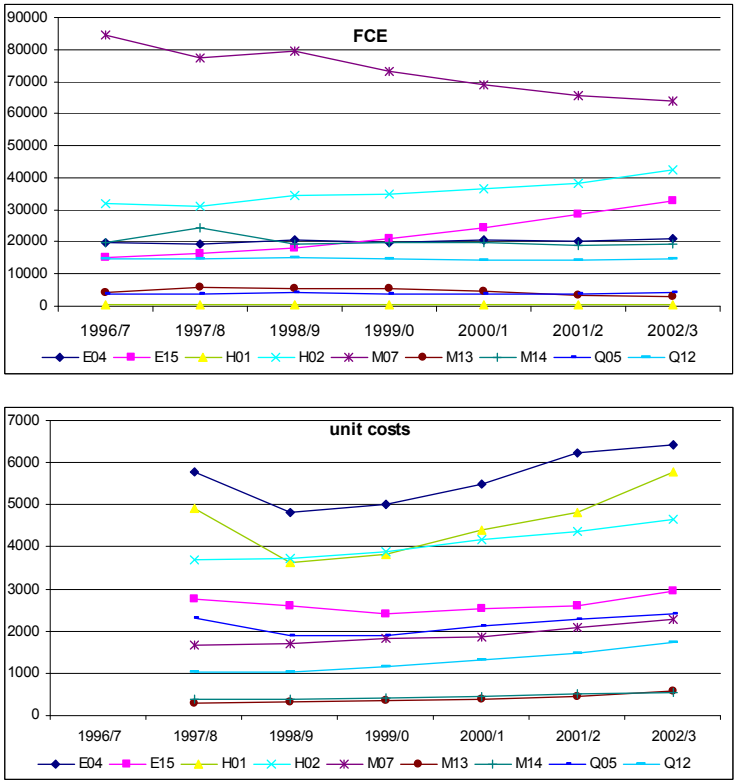
- The number of finished consultant episodes (y_{jt}), median waiting time (w_{jt}) and mortality rate (m_{jt}) were extracted from the Hospital Episode Statistics for each year from 1996/97 to 2002/03, where j indexes the HRG and t index the year. By incorporating data on w and m that are specific to each HRG, rather than aggregate measures, the index will be more sensitive to changes in the level and mix of activity. We also measured the mean age and proportion of female patients as initial indicators of patient mix.
- Costs for each HRG (c_{jt}) were derived from the Reference Cost Schedules from 1999/99 to 2002/03 as the activity-weighted average of the costs reported separately for day case, elective inpatient and emergency activity.
- Only a single temporal estimate was available from the published literature of the change in health status for each HRG (q_j), as measured by EQ-5D. To calculate QALY gain we assume that the reported health changes last for one year only. Of course, this is an unrealistic assumption for most treatments and would need to be addressed if applying the technique generally.

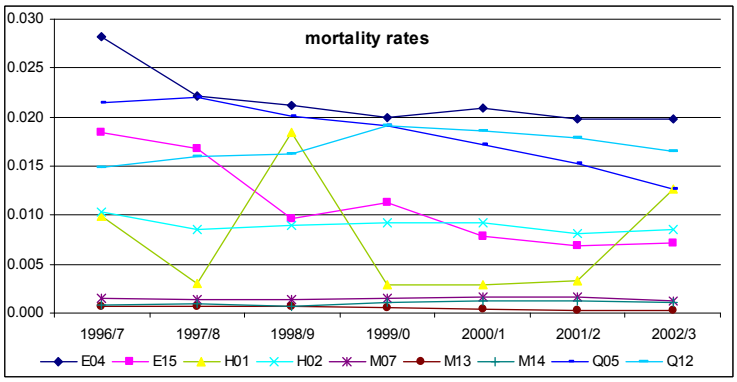
Table 3.6 Selected healthcare resource groups (2002/3)

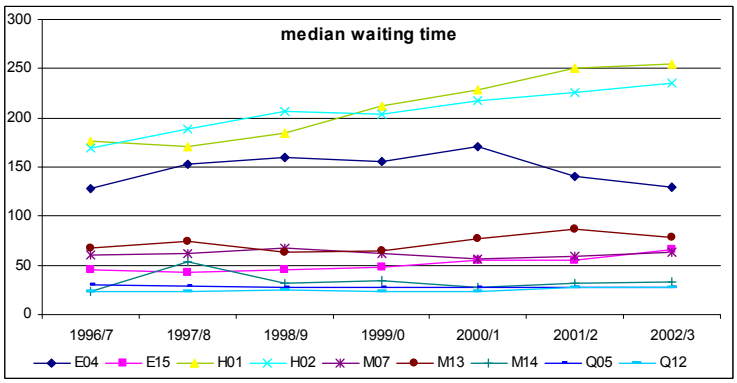
HRG	DESCRIPTION	FCEs	Change in EQ-5D	Unit Cost	Median Wait (days)	Mortality Rate *	Mean Age	% Male
E04	Coronary Bypass	21,033	0.06	£6,436	130	1.98%	64	80%
E15	Percutaneous Transluminal Coronary Angioplasty (PTCA)	32,982	0.16	£2,959	66	0.72%	62	73%
H01	Bilateral Primary Hip Replacement	316	0.42	£5,791	254.5	1.27%	62	45%
H02	Primary Hip Replacement	42,354	0.42	£4,672	235	0.85%	68	39%
M07	Upper Genital Tract Major Procedures	63,925	0.18	£2,267	63	0.12%	45	0%
M13	Non-Surgical Treatment of Genital Prolapse or Incontinence	3,093	0.04	£592	79	0.03%	57	0%
M14	Non-Surgical Treatment of Fibroids, Menstrual Disorders, or Endometriosis	19,304	0.04	£543	33	0.10%	34	0%
Q05	Extracranial or Upper Limb Arterial Surgery	4,029	0.04	£2,412	28	1.27%	67	62%
Q12	Therapeutic Endovascular Procedures	14,749	0.10	£1,729	28	1.65%	66	58%

Time series data on FCEs, weighted reference costs, median waiting times and mortality rates for each HRG are in Figure 3.8.

Figure 3.8 Selected health resource groups: time series







The data were used to construct two sets of indices:

1. A Laspeyre cost weighted index, CWAI, in which activities are valued using weighted Reference Costs (c_{jt}).
2. A Laspeyre QALY gain weighted index, QWAI, in which cost-weights are substituted for the change in EQ-5D score (q_j).

These indices were then adjusted to incorporate changes in either median weighting times (w_{jt}) or mortality (m_{jt}). The construction of the indices is shown in Table 3.7.

Table 3.7 Cost and QALY gain weighted activity indices

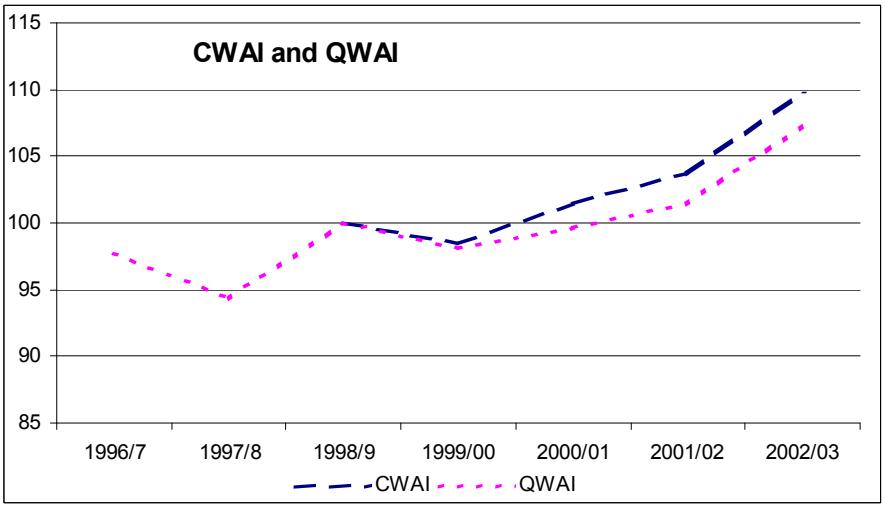
	CWAI	QWAI
Base-weighted	$\frac{\sum_j x_{jt} c_{j0}}{\sum_j x_{j0} c_{j0}}$	$\frac{\sum_j x_{jt} q_j}{\sum_j x_{j0} q_j}$
Adjusted by waiting time	$\frac{\sum_j x_{jt} \frac{w_{j0}}{w_{jt}} c_{j0}}{\sum_j x_{j0} c_{j0}}$	$\frac{\sum_j x_{jt} \frac{w_{j0}}{w_{jt}} q_j}{\sum_j x_{j0} q_j}$
Adjusted by mortality rate	$\frac{\sum_j x_{jt} \frac{m_{j0}}{m_{jt}} c_{j0}}{\sum_j x_{j0} c_{j0}}$	$\frac{\sum_j x_{jt} \frac{m_{j0}}{m_{jt}} q_j}{\sum_j x_{j0} q_j}$

3.3.5.2 Results

A core issue is whether estimates of productivity change are sensitive to whether activities are weighted by cost or by QALY gain. If the ratio of unit costs to QALY gain is the same across all outputs at all points in time the two sets of weights yield the same index of NHS output. Table 3.6 shows that Coronary Bypass (E04) would receive a relatively high weight in cost-based index but less so in a QALY-based index, where the change in EQ-5D amounts to 0.06. In contrast, hip replacements and upper genital tract procedures receive greater weight in the QALY-based index.

Figure 3.9 shows that the CWAI and QWAI measures both indicate increases in output for this small set of HRGs but with a faster rate of growth with the CWAI measure.

Figure 3.9 Comparison of base-weighted CWAI and QWAI



The CWAI index is more sensitive to the mortality adjustment than to the waiting time adjustment and the three versions of CWAI in Figure 3.10 follow a broadly similar temporal pattern. The QWAI measure is more sensitive to waiting time and mortality adjustments than the CWAI measure (Figure 3.11) and the index adjusted by waiting time suggests a reduction in output in contrast to the simple QWAI and the mortality adjusted QWAI.

Figure 3.10 Adjusted CWAI Index

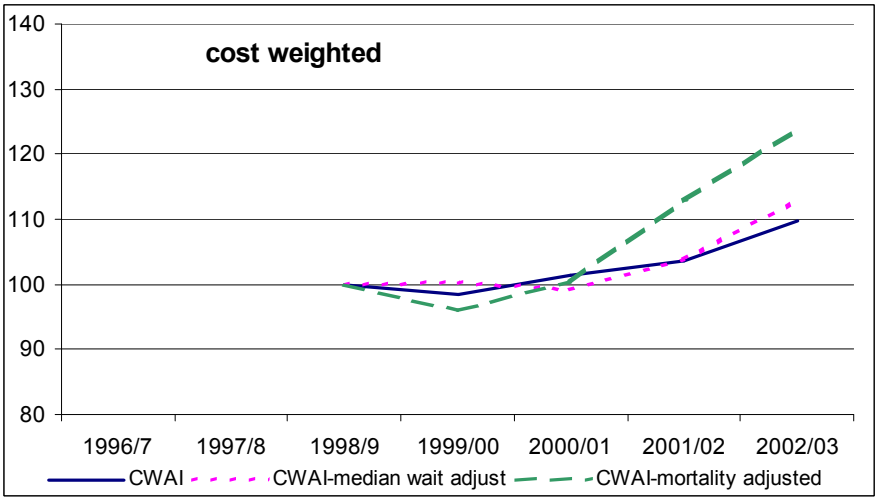
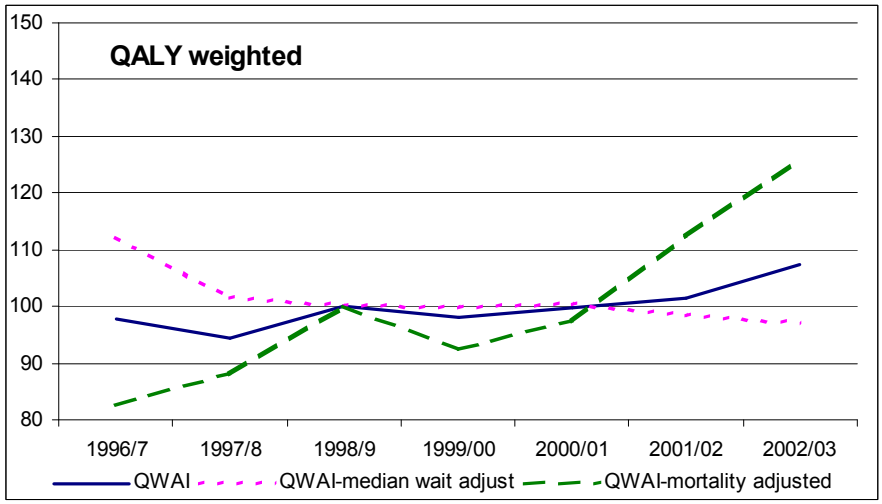


Figure 3.11 Adjusted QWAI Index

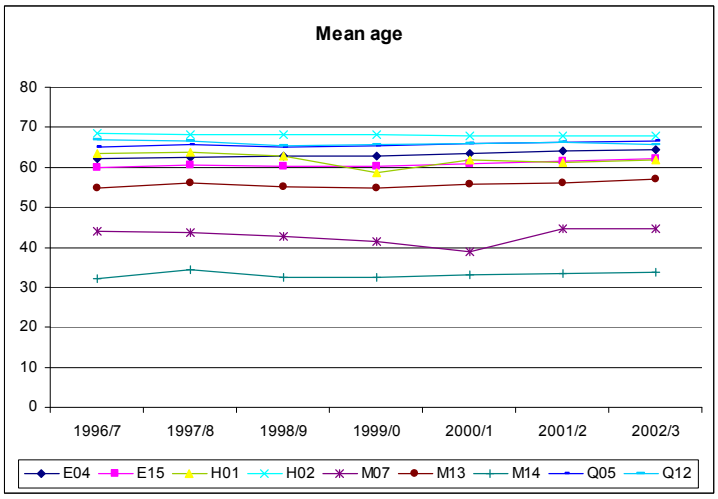


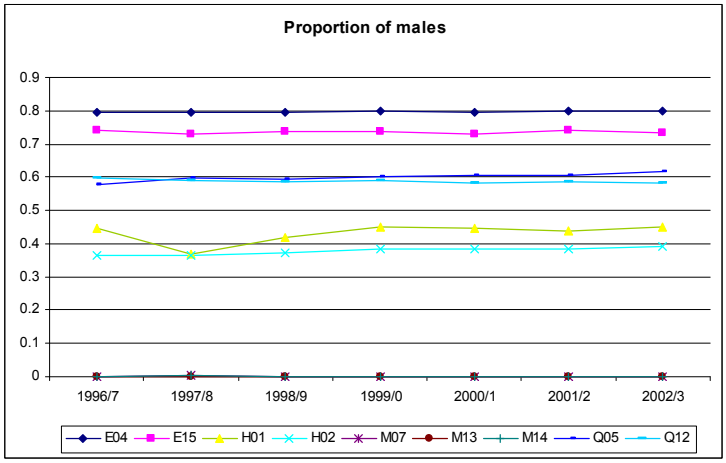
3.3.5.3 Further issues

The other outcome proxies are incorporated in the indices in a very simple ad hoc way which makes very strong assumptions about their relative values and their relationship with the volume of output. It is for example not obvious that a 10% reduction in waiting time for an HRG is equivalent to a 10% increase in the value of output. Similarly one might regard a 10% increase in mortality differently depending on the absolute mortality risk associated with an HRG. Further, if the QALY gains are based on the discounted difference in the health stream over time it would not be necessary to incorporate both mortality weights and QALY gains as we do in one of the indices. We will be investigating other weighting schemes. The issue is discussed further in section 5.2.

The QALY gain (and other outcomes) from an HRG depend on the mix of patients. The time series of mean age and proportion of female patients in Figure 3.12 do not show much temporal variation but of course these are crude indicators of patient mix. We will be investigating other possible indicators, such as more elaborate measures of demographic mix and the use of attributed socio-economic indicators from the 2001 Census.

Figure 3.12 Patient mix





3.4 Non-health outcomes

3.4.1 Waiting times

Data on changes in in-patient waiting time are routinely available from HES statistics at the procedure, trust and national level. Data on outpatient waiting times for first outpatient appointment are available from quarterly trust returns at the specialty, trust and national level. Data on waiting time for second or subsequent outpatient appointments are not collected by the DH. As waiting time for these appointments may be related to obtaining diagnostic tests, the missing data is regrettable especially as part of new investment in the NHS is to improve performance in this area. Changes in the choice and certainty of date of treatment can be tracked by changes in

the number of booked admissions and data on changes in the rate of cancelled admissions. The Audit Commission casts some doubt on the value of the way the DH measures cancellations (Audit Commission, 2003) but in principle, routine data are available that might be used to indicate improvement or deterioration in delivering this attribute of health care.

3.4.2 Other non-health outcomes

Other non-health outcomes pose more serious data problems. A main source of UK data are the annual surveys of patient experience. However, there are important problems with using surveys of patient satisfaction to obtain evidence of changes in the quality of care delivered by the NHS. Patient satisfaction is strongly related to expectations: "Satisfaction only implies that expectations have been met" (Cleary and Edgman-Levitan, 1997). The patient with low expectations registers as a satisfied patient. A US study of the determinants of patient satisfaction (Jackson, Chamberlin and Kroenke, 2001) identified a number of factors that may limit the value of using patient satisfaction data in the measurement of NHS productivity. Health state, as measured by instruments such as EQ5D or SF-36 is a predictor of patient satisfaction. The less healthy are less likely to be satisfied by their treatment. The NHS is pursuing a policy of reducing the use of ineffective treatments and of treatments that are not cost-effective. However, this could be reflected in reduced patient satisfaction. Analysis of the determinants of satisfaction indicated that satisfaction increases when patients are offered diagnostic tests (e.g. screening), prescriptions or procedures even though the scientific evidence suggests these activities are of little clinical value. Withdrawing these procedures in the interests of improving the quality of care could lead to increases in reported patient dissatisfaction. Patient satisfaction may not be correlated with the delivery of better quality treatment. Patients surveyed in settings (GP surgeries, hospitals) that are regularly used "typically report being satisfied 90% of the time". The NHS reports 85% patient satisfaction (Department of Health, 2004). Given the high proportion of patients reporting they are satisfied, the scope for obtaining improvements in patient satisfaction for a productivity growth exercise may be limited.

Picker Europe has worked with the Department of Health to develop instruments for

monitoring the NHS patient experience and trusts are now required to undertake annual surveys as part of CHI performance monitoring. To reduce some of the problems identified in the literature on expectations, patients are asked to report on what did or did not happen to them in the course of treatment. An analysis of data from the 1999 CHD patient survey (Jenkinson *et al.*, 2003) provides some insight into the determinants of the patient experience in the NHS. Older patients, men and patients from lower socio-economic classes report better experience of treatment. While the impact of age and gender on overall score is small, it does affect the ranking of trusts. Of more importance for the work on productivity, only 5% of the variance in the patient experience index score could be accounted for by trust level factors. Data from surveys of the patient experience may have some use in tracking national trends in non-health attributes of the NHS but are likely to be of limited value when disaggregated to the trust level.

If it is possible to obtain the raw data from these annual surveys, it may be that analysis of a subset of questions could be used to monitor changes in attributes such as nursing care, communication, food and the environment. In section 3.6 we suggest a means of using these data which may be preferable to ignoring them.

3.5 Valuation of outcomes

3.5.1 Health gains

The National Co-ordinating Centre for Research Methodology (www.publichealth.bham.ac.uk/nccrm) has funded a programme of research on measuring the value of health gains and the value of QALYs. The results of the research will not be available within the limited time scale of our project, but could be highly relevant to any future DH application of output and productivity methodologies developed in this project. However the research briefs for the initiative do not cover changes over time in the monetary value of health. Since there is some theoretical argument and a little empirical evidence that the value of health grows over time, TFP growth may be underestimated if a constant monetary value of a QALY is assumed (Gravelle and Smith, 2001).

Other possible sources for a monetary QALY value are to examine the valuation implied by public sector decisions. For example NICE seems to operating with a cost per QALY threshold of £30000. Other possibilities are to use the explicit valuations by the DETR of lives saved in transport projects and to translate these into a value per QALY.

3.5.2 Non-health outcomes

In the UK the market for private health insurance provides evidence for some non-health attributes. There is no reason to believe the improvement in health will differ as between NHS and private treatment but, with private health care, waiting time is reduced and facilities are of a higher quality. Individual willingness to pay for private health insurance or for private treatment gives an indirect valuation of these non-health attributes of health care (Besley, Hall and Preston, 1999). However, patients who purchase private health insurance tend to have higher incomes than the population of NHS patients and the evidence relates primarily to demand for elective surgical procedures.

An alternative source of data on willingness to pay to reduce waiting times is evidence from discrete choice experiments. A recent review of the literature (Ryan, Odejar and Napper, 2004) reported that few studies addressed the issue of the monetary value of reducing waiting times for health care and contrasted this with the significantly greater volume of work on the value of time saving in transport. Of the six UK papers, only one sampled the English population. The other five were location or procedure specific. Ryan summarises the available evidence converting to 2002/03 prices. Propper's analysis of English data suggests estimated values between £36.25 and £94.19 for a one month reduction in waiting time. Hurst's study of waiting time for non-urgent rheumatology estimated values between £11.95 and £23.68 per week. Ryan points out that these two studies give similar values assuming a linear additive model.

A large scale discrete choice experiment focused on waiting time is currently being carried out as part of the evaluation of the London Patient Choice Project (Burge *et al.*, 2004) but it is unlikely to yield data on the monetary valuation of reductions in

waiting time. It may however provide evidence on the tradeoff between waiting time and other attributes, such as distance, and if these other attributes can be valued, an implied valuation of waiting time can be derived.

The view that individuals have preferences over standards of accommodation (room size, privacy, environment) is commonplace for economists who study the demand for housing. Techniques for estimating hedonic prices have long been used to estimate market values for the various characteristics of houses bundled into a single market exchange. It is also commonplace that over time the relative valuation of characteristics can change as real incomes increase. In the absence of market prices for hospital services, stated preference techniques would be required or values imputed from other sectors.

Patients bear costs in accessing and using the NHS and NHS policy changes can alter these costs, for example by the location of facilities or by introducing new ways of accessing the NHS via NHS Direct or Walk In Centres. Such changes can be regarded equivalently as quality improvements or as cost reductions. In the private sector such costs would be reflected in the prices paid for goods. In the NHS although there are no money prices for final consumption, choice of provider may reveal the marginal rates of substitution between say distance, waiting time and other outcomes.

3.6 Marginal cost as a guide to valuation of non-health outputs

If we can find measures of the non-health outcomes, but not their marginal social values, then we can estimate a cost function for the production of outcomes and use the marginal cost of each outcome, except for the QALY gain for which we have a price, as a measure of its marginal social value. The marginal cost measure, may for the reasons we have already noted in section 2.3, not reflect the marginal social value but it is at least a defensible approximation, and certainly better than ignoring an outcome: a marginal social value of zero is probably a worse approximation.

We will investigate the feasibility of estimating the marginal cost of providing ancillary outcomes by estimating a cost function using NHS Trust level data. We denote the outcomes by a vector **q** where q_{ij} is the amount of outcome *i* per patient in

trust j. The cost function explains the operating costs, C_j in terms of the logs of the various outcomes. Since the costs of operating a hospital are bound to depend on the nature of treatment offered, it is necessary to represent not the total number of QALYs per patient in **q** but the number of QALYs per patient in each of a range of treatments. In practice it may not be possible to distinguish the QALY output of different trusts for the same treatment, so instead of the QALY output for each major treatment it may be necessary to include simply a number of variables representing the proportions of patients receiving each of a number of major treatments with a residual category to deal with minor treatments. We do, however also include quantitative indicators which either represent the output per patient of ancillary services directly or the logs of quantitative indicators produced from qualitative data. Thus the cost function might be

$$\ln C_i = \sum_j \eta_j \ln q_{ij} + \gamma + \varepsilon_i \tag{23}$$

This can be estimated either by least squares regression or stochastic frontier analysis. The latter method has the advantage that it takes account of the fact that trusts are more likely to be inefficient (producing at above best-practice cost) than super-efficient (producing at below best-practice cost).

The above equation of course includes fixed effects. These can be removed by looking at the change in costs between two different years to give an equation

$$\Delta \ln C_i = \sum_j \eta_j \Delta \ln q_{ij} + \gamma^* + \varepsilon_i \tag{24}$$

It should be noted that γ^* represents the percentage change in costs not attributed to any increase in output. If there were no other data this could be regarded as the increase in the log of the price index while $\sum_j \eta_j \Delta \ln q_{ij}$ gives the quantity index increase. In practice, however, having decided to use this method to deal only with those outcomes which cannot be valued in any other way we prefer a price index defined with reference to our preferred quantity index.

It should be noted that we do not impose any adding up restriction on the η_j. It might

seem appealing to impose a homogeneity restriction that the components should add to one. But the ancillary indicators are often defined using arbitrary scales (e.g. with patient satisfaction) and allowing η_j to be freely determined makes whatever exponential transformation is needed to the original data to align them to costs. Thus we can be relaxed about the units in which the q_{ij} are measured.

We can derive the price associated with the quantity q_j by noting that the marginal cost of one extra unit of q_j is given as

$$\frac{\partial C}{\partial q_j} = \pi_j = \frac{\eta_j C}{q_j};$$

π_j is the price to be attributed to outcome j in working out a base-weighted index of output. We would probably not want to stick too strongly to the functional form implied by the cost function but might instead prefer to re-estimate as necessary in order to obtain up to date prices.

Note that there may be some variables which are politically important but which seem to have a zero cost because low-cost trusts do them as well a high cost trusts. If patient satisfaction costs nothing to produce then there is a good argument that it should have a zero weight in an output index, like politeness in GDP. Thus the fact that the regression might not "work" is not a good argument against its use.

3.7 Composite measures of health sector output

The development of composite indicators may be relevant to the analysis of productivity in healthcare because composite indicators are essentially a weighted aggregated measure of healthcare outputs (performance indicators). The methodology used to construct a composite measure of healthcare outputs requires the specification of a set of weights to combine the underlying dimensions which reflect social valuations. There has been much debate in the literature as to the best ways of eliciting such weights and evidence suggests that the outcome on the composite is highly sensitive to these methodological choices (Jacobs, Smith and Goddard, 2003). Examples of the development and use of composite measures of healthcare output are described in more detail in Appendix C.

4 Inputs and growth accounting

4.1 The index number approach to measuring productivity

The starting point for estimating productivity growth is the Divisia index, given in equation (2) of section 2.2. In practice however data are available only in discrete intervals so that implementation requires the specification of a particular functional form for the production function. Although simple specifications such as the Cobb-Douglas were used initially in research on productivity growth, it is now common practice to employ a 'flexible functional form' for the production function. Flexible functional forms are those that satisfy the property that they are second order approximations to any arbitrary production function. Jorgenson and Griliches (1967) based their analysis of TFP measurement on the Törnqvist discrete approximation to the Divisia index and this approach has become standard since then. Later Diewert (1976) established an important link between practice and theory by showing that the Törnqvist discrete approximation to the Divisia index, is, in fact, an exact index number if the production function has the translog functional form. An overview of the development of the index number to calculate TFP in theory and practice is presented in Hulten (2001).

For simplicity in this section we assume only one output (or the existence of an aggregate output using the method outlined in section 2.2), Y, and drop the subscript i for firm i. It is common to divide inputs into three groups, capital, K, labour, L, and intermediate input, I. Then comparing periods t and t-1, and rearranging terms, the Törnqvist total factor productivity index is given by:

$$\ln TFP_t - \ln TFP_{t-1} = (\ln Y_t - \ln Y_{t-1}) - \varpi^K (\ln K_t - \ln K_{t-1}) \tag{25}$$

$$- \varpi^L (\ln L_t - \ln L_{t-1}) - \varpi^I (\ln I_t - \ln I_{t-1})$$

where K, L and I are aggregate inputs of capital, labour and intermediates and ϖ is the share of each aggregate input in the value of output, averaged over period t and t-1. For example for labour input, this is given by:

(26) $$\overline{\omega}^{L} = 0.5\left(\frac{w_{t-1}L_{t-1}}{p_{t-1}Y_{t-1}} + \frac{w_t L_t}{p_t Y_t}\right)$$

with analogous expressions for capital and intermediates. In practice, when output prices are unavailable, the denominator in the above expression is replaced by total expenditure on inputs (total costs) and so the weights are average cost shares.

The earlier measures of TFP based on Solow (1957) showed very large contributions from the residual TFP term. The seminal contribution of Jorgenson and Griliches (1967) was to show that careful measurement of the variables included in the TFP equation led to a reduction in the contribution of the residual TFP to output growth. The use of aggregates for capital, labour and intermediate inputs missed important quality changes whereby higher quality inputs were substituted for lower quality ones, although generally at higher cost. Hence some part of measured TFP was in fact due to greater quality adjusted input growth.

The essence of the Jorgenson and Griliches method was to replace aggregate inputs by Törnqvist indices of their components. This approach distinguishes *k* types of capital goods, *l* types of labour and *f* types of intermediate goods, with each aggregate input measured by:

(27) $$\ln K_t - \ln K_{t-1} = \sum_k \overline{\omega}^{kK} (\ln K_{kt} - \ln K_{kt-1})$$

(28) $$\ln L_t - \ln L_{t-1} = \sum_l \overline{\omega}^{lL} (\ln L_{lt} - \ln L_{lt-1})$$

(29) $$\ln I_t - \ln I_{t-1} = \sum_f \overline{\omega}^{fI} (\ln I_{ft} - \ln I_{ft-1})$$

where $\overline{\omega}^{kK}$ is the share of type *k* capital in the total value of capital, again averaged across the two time periods, and similarly for *L* and *I*.

To see how this disaggregation incorporates input quality changes, consider labour input L and assume there are two types, more skilled (doctors - d) and less skilled (nurses - n). Since wages measure marginal products we expect that $w^d > w^n$. Then changes in labour quality can be measured as:

$$(30) \quad \left[\overline{\omega}^{dL}\left\{\ln\left(L_t^d\right)-\ln\left(L_{t-1}^d\right)\right\}+\overline{\omega}^{nL}\left\{\ln\left(L_t^n\right)-\ln\left(L_{t-1}^n\right)\right\}\right] - \left[\ln\left(L_t^d+L_t^n\right)-\ln\left(L_{t-1}^d+L_{t-1}^n\right)\right]$$

The second expression is approximately equal to the employment share weighted average of the growth rates in L^d and L^n. Given the higher wages of type d labour, if the growth rate of L^d is greater than L^n then labour quality is increasing. Similarly labour quality can decline if low cost workers are substituted for higher paid ones. Obviously if labour growth is the same for the two types, equal to the growth in aggregate labour, then labour quality growth is zero regardless of the relative wage rates – this follows from the fact that the wage bill shares in the above equation sum to one. In addition labour quality will be influenced by the rate at which wages are changing. If wage rates also increase by more for doctors than nurses between t and t-1 then the growth in labour quality is higher than in the constant wage case. However since the shares are averages across two adjacent periods differences in wage growth tend to be small relative to the impact of changes in labour input.

In order to capture changes in labour quality in their entirety it would be necessary to capture all changes that impact on the effectiveness of labour input, in particular changes in the skills of the labour employed. This may be difficult to achieve in practice since it requires splitting the labour force into occupation and skill levels as well as other demographic factors such as age that may affect labour quality. The division of labour input in the NHS is discussed in section 4.4.1 below.

An important property of the above approach is that input weights vary through time. Fixed base indices are likely to give biased results if weights are fixed for long periods of time. Suppose a new type of drug is substituted for an old less effective one. Base period weights would give too high a weight to the older drug and too low to the new drug, so underestimating the change in material inputs. Correctly

measuring the importance of changes in input use requires information both on their growth rates and changes in the importance of these inputs in total expenditure. Chain linking has now become standard in most countries national accounts.

The index number structure we have proposed is focused firmly on giving good measures of volume movements of inputs, outputs and productivity. The use of Törnqvist indices does, however, have the implication that there is no analogous formula for price changes, since we need a volume index multiplied by a price index to give an index of value of inputs/outputs relative to the base period.

We do not see this as a problem. Price indices, where they are required, can be calculated as purely derived variables, found on dividing value indices by the volume indices evaluated as we suggest. We regard this as more satisfactory than the alternative of compromising over the choice of volume index so that there can be a symmetrically defined price index.

4.2 Attribution of productivity growth to sectors

Economists are typically concerned both with measuring the outputs of particular industries and measuring flows of goods and services to final demand. Productivity and productivity change is usually regarded as an industry characteristic rather than something associated with final demand and is measured on an industry basis. If one is concerned solely with measuring overall economic activity then measurement in terms of final demand is perfectly adequate. Sales from one industry as inputs to another net out; the only intermediate inputs remaining are imports which are treated as negative final demand in the GDP identity.

However, we expect, in constant prices, to be able to measure GDP using output indicators representing (possibly imperfectly) the net value added of each industry. Chain-linked growth in the expenditure measure of GDP should equal chain-linked growth in the output measure of GDP.

Productivity measures compare growth in outputs with growth in factor and material inputs. It is not possible to allocate factor inputs directly to categories of final demand

except through the mechanism of an input-output table. However, we do of course know the factor input supplied to the economy as a whole. This means that we can identify productivity growth in the economy as a whole even though we cannot identify productivity growth associated with any particular component of demand.

The distinction between industry and final demand may be of some importance for assessing the health service. The reason is, of course, that the health "industry" is a heavy user of drugs produced by the pharmaceutical industry which is part of manufacturing. Careful measurement of the consumption of health services combined with rough and ready measurement of the output of pharmaceuticals means that output changes might be attributed to the NHS when in fact they are the consequences of improvements in manufacturing.

Suppose a new more effective drug becomes available, which also costs more. Pharmaceutical input is typically measured by the value of expenditure on drugs, deflated by a price index. If this price index does not take account of the increased effectiveness of the drug then the growth in the real volume of drug input could be underestimated. As TFP is measured as a residual, this will lead to a corresponding overstatement of TFP growth in the health sector. This is a common problem in index numbers which use a residual formula since all measurement error is subsumed in the residual. In practice this is not a trivial issue, e.g. Lichtenberg and Virabhak (2002) present evidence for the US that people who used newer drugs had better post-treatment health than people using older drugs for the same condition, controlling for a range of other influences such as pre-treatment health, age, sex, etc.

Suppose now that the use of the new drug also saves on some other input, say labour. Since labour input is now falling and, assuming no changes in wage rates, the share of labour is also falling, aggregate input growth will be less than before and TFP growth higher. A question then arises as to whether this additional saving of labour input and corresponding rise in TFP should be seen as arising from the health or the pharmaceutical sectors.

Conventional practice allocates growth in TFP attributable to substitution of drugs for other inputs to the NHS, although the change would not have occurred if the new drug

had not been developed. This change can be interpreted as picking up organisational changes and adoption effects that lead to more effective use of the new drug. Thus it is picking up the influence of the information in the hands of doctors that influences rates of adoption of the new treatment and the willingness to reorganise practice to make the most of the new drug. In practice it is difficult to separate these effects since doing so requires a measure of the impacts of the new drug on the health of the patient when other inputs have not been changed, an event that is not observed in practice.
One possible way around this problem is to employ a method such as a hedonic regression to quality adjust the drugs deflator. Then real drug input growth would rise and there would be a corresponding fall in TFP growth in the NHS. Without the quality adjustment to prices, changes originating in another sector are incorrectly attributed to the NHS.

This does not matter if we are concerned only with measuring the amount of health supplied to final demand and the political debate is likely to focus on this. But it does matter if targets are defined in terms of the productivity of the health service as they have been by past public service agreements. Failure to take proper account of improvements in the pharmaceuticals industry raises the prospect that health service managers might be credited with the benefits derived from the work of research chemists. If we wish to measure properly improvements in the provision of health care we need to be aware of their source.

The method outlined in the previous section has been used extensively to compare TFP growth rates between sectors of the economy (Jorgenson, Gollop and Fraumeni, 1987; O'Mahony, 1999; Triplett and Bosworth, 2002). However it is also useful as a device to investigate the relative contribution of different inputs to output changes. Hence the approach is termed growth accounting. Thus the focus is not only on output growth but also on the extent to which output growth is driven by changes in input quantities and qualities. From the perspective of the NHS there may be interest in knowing how far improvements in health outcomes are due to the use of better quality inputs such as skilled labour, high technology capital or better drugs, even if some of this reflects technological changes originating elsewhere.

4.3 Productivity and efficiency

The growth accounting methodology outlined above divides output change into that due to the change in inputs and residual TFP growth. TFP growth is often equated with the rate of technical change. This, however, assumes that there is no inefficiency in the health service (or if there is, that it is unchanging). If there is inefficiency in the provision of services, output may increase without either an increase in inputs or technical change. This will occur if the provider becomes more efficient. To see this, consider the following case, where for simplicity there is a single output (Y), which is produced by a vector of n inputs ($\mathbf{z}$). The production technology is

$$Y = f(\mathbf{z}, \mathrm{T}; \boldsymbol{\beta}) \cdot \mathrm{E} \tag{31}$$

where $\boldsymbol{\beta}$ is the technology parameter vector to be estimated[1] and E is what Farrell (1957) calls 'technical inefficiency' (the ability of a unit to obtain maximal output from a given set of inputs). From section 2 we know that:

$$TFPG = \frac{\dot{y}}{y} - \sum_n \varpi_n \frac{\dot{z}_n}{z_n} = \frac{f_t}{f} + \frac{\dot{E}}{E} \tag{32}$$

This tells us that the growth in TFP is the sum of technical change (f_t/f) and technical efficiency change ($\dot{E}/E$). More generally, efficiency could also include what Farrell (1957) calls 'allocative inefficiency'. This refers to the ability of a unit to use the inputs in optimal proportions, given their respective prices.

4.4 Measurement of inputs

The measurement of inputs requires information on both their quantity and quality, for the conventional three broad categories, labour, capital and materials. This section considers measurement issues and an initial overview of data availability. First, each of the three broad categories are considered in turn, outlining the desired data requirements and discussing information readily available. Data on inputs are available for both the whole health sector as well as at more disaggregated levels, such as NHS trusts. Technological change in inputs, in particular in high technology

[1] This could be estimated using either econometric (stochastic frontier) or linear-programming (data envelopment or analysis) methods. For more on these see Appendix B.

equipment and drugs, can be important sources of quality change and so are discussed in a subsection on input price deflators. Finally we consider unpriced inputs, including patient time and effort, which also play an important role in the NHS. There are some gaps in our knowledge on sources of input data and in some cases there will be a need to choose between sources. To ensure the most accurate estimates in the final report, the researchers will seek advice from experts, including those at ONS and the Department of Health, at all stages of the calculations of input indices.

4.4.1 Labour input

Measuring aggregate labour input requires data on the number of workers employed and average annual hours per person employed. Since the employment of different types of workers, e.g. workers distinguished by their skill level, will have different implications for productivity, it is necessary to adjust for labour quality using the method outlined in equation (28) above. This in turn requires data on the proportion of each type in the total and their relative wage rates. In addition the total wage bill is required to estimate labour's share of total expenditure. This should include both earnings and non-wage labour costs such as national insurance payments and pension contributions.

In terms of labour input there are plentiful data available but no one source yields all the information required. The starting point for estimating total numbers employed will be the annual Department of Health Workforce Census which provides information on headcounts and labour input on a whole time equivalent (WTE) basis. This source underlies the ONS headcount measure as part of the estimate of UK public sector jobs. However there is some question over the reliability of this source as an accurate measure of headcount so additional sources will be required. For example additional information can be gained from the population weighted Labour Force Survey (LFS) which gives annual data on employment and hours worked for industry SIC group 85.1, 'human health activities'. This distinguishes between hospital, medical practice, dental practice and other activities and whether the workers are employed by the private or public sectors. Another possible source of information is Census 2001, which also contains data on employment by industry and can be a useful check on the LFS and the DH workforce census. The LFS is the only annual

source of information on average hours worked with again the 2001 population providing a useful check.

Tracing the changes over time in the utilisation of various types of workers will be an important addition to using a simple headcount. The DH workforce census also shows headcounts and WTEs for a large number of different types of workers including doctors, nurses, associated health professionals (physiotherapists, dieticians etc.) and community health professionals. The annual NHS GP census provides data on types and WTE numbers of GPs.

Estimating an aggregate index of the change in the quality of the workforce, requires information on the relative wage rates of categories of workers. One possibility is to attempt to match the DH census data with the Trust Financial Returns (TFR) for the hospital sector. This distinguishes managers, medical staff, dental staff, nursing, midwifery and health visiting staff, scientific, therapeutic and technical staff and other auxilliary workers (see Appendix E). The Labour Force Survey will also be a useful source of information, in particular outside the hospital sector. For example there appears to be no routine DH source on GP remuneration which distinguishes between income and practice expenses. Since the LFS distinguishes hospital activities from medical practice activities and distinguishes numerous occupational categories, in principal it is possible to derive an estimate of GP remuneration from this source. However the LFS needs to be treated with caution as the sample sizes can be very small and so it may be necessary to average across a number of years' surveys. Finally data on salary expenditure and WTEs for all Trusts are also available in the CIPFA database so that it may be possible to match earnings and numbers of workers from this source.

Occupational classifications may miss some quality improvements if there are changes over time in the skills acquired by staff, for example the much more common use of nurses with degrees now than in the past. The Labour Force Survey contains some information on qualifications, distinguishing for example higher degrees, primary degrees and HE qualifications and so will be a useful source in tracing changes in the proportions of workers with varying across time. The LFS panel since 1993 could be used to estimate the impact of qualifications on earnings using

variations of the standard Mincer equations as used in O'Mahony and Stevens (2003).

We will also explore whether there are more detailed sources of information available indicating staffing by grade of staff and payroll costs by grade of staff. It is possible that such data may exist at a national level (e.g. from the Royal Colleges).

In all, it appears that there are numerous sources of information for measuring labour input. While the existence of so many sources is advantageous, considerable care will be required to ensure consistency and to produce a reliable estimate. This will be especially important with regard to examining quality change over time.

4.4.2 Intermediate input

Perhaps the most difficult of the three inputs to measure is real material inputs. This requires both expenditures and price deflators. Intermediate input is calculated by dividing into product areas, and deflating each by the most appropriate industry producer price index – a further discussion of price indices is in section 4.4.5.

Table 4.1 shows such a decomposition of purchases by NHS Trusts derived from the Trust Financial Returns (TFR3); details on items included in TFR3 are shown in Appendix E. This highlights the importance of purchases from the pharmaceutical sector, and its increasing share over time.

Table 4.1 Shares of nominal intermediate inputs: NHS Trusts

	Drugs	Other clinical supplies	Stationery, postage, telephone	Other products	Energy, fuel and water	Hotel services	Other services
1997/98	28.4	4.9	8.1	3.5	7.2	11.6	36.4
1998/99	27.5	4.7	7.4	2.9	6.2	7.3	44.1
1999/00	26.7	4.6	6.7	2.6	5.4	10.8	43.2
2000/01	32.3	6.0	8.0	3.1	5.9	12.4	32.4
2001/02	34.5	6.3	7.6	2.7	6.2	12.2	30.5

Source: TFR3, Department of Health

It will be necessary to also gather information on intermediate inputs from Primary Care Trusts.

4.4.3 Capital input

Conventional measures of capital stocks use the perpetual inventory method that accumulates investment over time and subtracts retirements/depreciation. This method requires a long time series on capital expenditures in constant prices and so should be linked to the most reliable data available in the National Accounts. However capital expenditures in the National Accounts are available for the entire post-war period only for SIC 85 which includes social services and veterinary services in addition to human health activities. These distinguish four asset types, structures, equipment, transport equipment and intangibles (largely computer software). Surveys of capital expenditure can be used also to distinguish computing equipment and communications equipment using the method outlined in O'Mahony and De Boer (2002). These data need to be adjusted to take out social and veterinary services and the private health sector. In addition it is important to ensure that expenditures are defined on a user rather than owner basis given the increased use of private finance initiatives. Adjustment for the additional sectors and for private health could be made using data on capital expenditure from the Annual Business Inquiries and those at the Trust level using the financial returns by type of expenditure (TFR3). Data availability for general practice premises and other capital assets will also be considered.

Table 4.2 below shows gross fixed capital formation (GFCF) at current prices, together with asset shares, for the period 1993 to 2001. The most notable feature of this table is that asset shares have changed considerably over time, which demonstrates the necessity of deriving a capital stock estimate. For most of the period, construction has accounted for by far the largest share of GFCF in health and social services. After falling in the mid-1990s to 26.7% of GFCF in 1995, the share accounted for by construction has increased dramatically, reaching 63.9% in 2001. 1995 was the only year during this period in which construction did not account for the largest share of GFCF, when medical instruments accounted for 40.1%. The share accounted for by medical instruments then fell in the following years, before increasing again in 2000. Computers have accounted for a declining share of GFCF over this period, falling from 8.7% in 1993 to 1.5% in 2001. However, as noted in section 4.4.5 declining prices for computing equipment complicate this picture so that real stocks of computing equipment might well be rising. Transport equipment has

also experienced a declining share over time, from 6.1% in 1993 to 1.9% in 2001. Software, at 3.7%, accounted for a lower share of GFCF in 2001 than in 1993, but has generally remained at approximately 3-4% throughout the period. The share of GFCF accounted for by telecommunications equipment has fallen slightly since the mid-1990s, although this has always accounted for the smallest share, standing at just 0.3% in 2001. Finally, the share of GFCF accounted for by other equipment has fallen considerably over the period, from 17.1% in 1993 to 3.8% in 2001.

Table 4.2 UK gross fixed capital formation, current prices

	1993	1994	1995	1996	1997	1998	1999	2000	2001
GFCF, £ million									
Construction	1117	971	743	1671	1791	2 215	2 441	2 307	2 193
Transport equipment	141	134	134	95	76	105	95	90	66
Medical instruments	337	696	1118	944	521	543	667	926	851
Computers	200	193	189	89	49	52	64	60	53
Telecommunications equipment	13	26	42	19	11	11	14	13	12
Software	110	104	111	90	89	126	148	138	127
Other equipment	392	408	450	202	114	130	155	147	132
Asset shares									
Construction	48.4	38.3	26.7	53.7	67.6	69.6	68.1	62.7	63.9
Transport equipment	6.1	5.3	4.8	3.1	2.9	3.3	2.7	2.4	1.9
Medical instruments	14.6	27.5	40.1	30.4	19.7	17.1	18.6	25.2	24.8
Computers	8.7	7.6	6.8	2.9	1.8	1.6	1.8	1.6	1.5
Telecommunications equipment	0.6	1.0	1.5	0.6	0.4	0.3	0.4	0.4	0.3
Software	4.8	4.1	4.0	2.9	3.4	4.0	4.1	3.7	3.7
Other equipment	17.0	16.1	16.1	6.5	4.3	4.1	4.3	4.0	3.8

In constructing capital input from investment data it is necessary to incorporate an assumption on rates of depreciation by asset type and then to aggregate across asset type. The theoretical method to do so was set out some years ago in Jorgenson *et al.* (1987) but there remain some practical considerations in applying this approach. These include the choice of depreciation rates and rates of return in constructing user costs of capital. A further complication is that the Treasury is undertaking an audit of all government assets at replacement cost, including the NHS. This will yield a benchmark estimate of capital stocks. Preliminary results suggest that this benchmark is not consistent with the stock estimates calculated using the perpetual inventory model. This finding is not unusual and was also found to be the case in comparing PIM and wealth stocks in countries such as the Netherlands and Japan. Thus it may be necessary to choose between the two methods. The research team will discuss alternatives with ONS.

4.4.4 Unpriced input

Patients bear costs in accessing and using the NHS and NHS policy changes can alter these costs, for example by the location of facilities or by introducing new ways of accessing the NHS via NHS Direct or Walk In Centres. Such changes can be regarded equivalently as quality improvements or as cost reductions. It is necessary to consider whether it is possible to quantify and value such changes, for example, one possibility may be to examine data on patient travel times to NHS facilities.

Note that another significant unpriced input into the NHS is unpaid care. The national accounting framework does not include unpaid work carried out in the home, so that following these guidelines also suggests this element should be excluded. Measurement of this poses a number of difficult issues, including problems with data availability and so is seen as beyond the scope of the current study.

4.4.5 Input deflators

To derive real material input requires relevant deflators, essentially dividing this component by type of input used and applying relevant deflators to each component. Some items such as electricity, fuel and business services can employ aggregate industry deflators. The main problem in terms of deflators is in calculating a suitable

series for pharmaceuticals. There is no price index or prices available for hospital drugs as they are all negotiated at discounts to list prices but price indices are available for primary care, calculated by the Department of Health. A careful analysis of existing data and how they might be combined to produce a reliable overall index will be an important contribution of the research project. At the very least we can use the producer price index for pharmaceuticals, possibly adjusting this to take out price changes in drugs sold directly to consumers. Additional information may be derived from import unit values.

There are also concerns related to defining appropriate deflators for capital assets. Medical care capital inputs include MRI scanners, which greatly improve diagnostics, as well as equipment that facilitates non-invasive surgery and improvements in patient monitoring. Much attention has been given in recent years to measuring "high tech" capital at the economy wide level, where the research focus is the contribution of high tech capital investment to economic growth and productivity improvement. Aggregate and industry-level studies for the US include Jorgenson and Stiroh (2000) and Triplett and Bosworth (2003). Comparisons of the relative contributions of high tech capital in European Union countries and the US appear in O'Mahony and van Ark (2003). The standard method to adjust for quality change in capital is to apply "hedonic" price indices to form investment deflators.

However nearly all this research literature concerns information and communication technology (ICT) equipment with little work carried out for other types of capital. Production of new and improved medical equipment is driven by the same technology that drives performance improvements in semiconductors and computers. The national accounts provide investment price deflators for aggregate equipment only but it may be possible to use producer price indices for medical equipment as an alternative.

Finally, as noted in section 2.12 there remains the issue of whether input prices measure marginal social costs. We believe that in some important cases input prices are a misleading means of aggregating inputs to construct measures of input use. Since relative input prices also affect decisions in the NHS on the best way of providing services we think that further work on the topic is merited. However given

our resource constraints we do not feel we can tackle the issues properly in this project. We therefore propose to adopt the conventional assumption that input prices measure marginal social costs and so are an appropriate price to use in constructing measures of input use in the NHS.

4.5 Non standard outputs and associated inputs

As discussed in section 2.8 there are a number of non standard outputs that are difficult to incorporate in the framework outlined in previous sections. Failure to account for these will have an impact on productivity since these activities will consume inputs without any corresponding measured increase in output. The impact on the final results may be non-trivial, if, for example, the NHS increases its expenditure on training.

Ideally the output measures would be adjusted to reflect these activities. In the case of training the incremental change to the lifetime earnings of health workers, both those who remain in the NHS and those employed in private health, due to NHS training could be used to measure training outputs. In practice estimating this is difficult since it requires data on earnings of persons undertaking training relative to those who receive none. The estimation problems are likely to be more severe for other non standard outputs such as R&D, public health, diagnostic testing and screening.

An alternative approach might be to estimate the costs of such activities and either add these to outputs or subtract from inputs. Some information does exist in previous studies on the costs of training, e.g. Netten and Curtis (2003). Such estimates could be employed to consider the sensitivity of our baseline estimates to incorporating non standard outputs. Nevertheless it is unlikely that the data required to undertake a comprehensive treatment of non standard outputs are readily available.

5 Conclusion and the way ahead

5.1 Introduction

Our main aim is to measure the aggregate outputs, outcomes and inputs of the health sector and thereby construct aggregate NHS level indicators of these and of productivity. The following summarises the methodology and the next steps required to move ahead from the methodological focus of this paper to implementation. For simplicity this summary describes the methodology in terms of base-weighted (Laspeyres) indices, but in the implementation stage the research may well use other index number formulae.

5.2 Outputs

5.2.1 Index structure

The key to this is the identification of the outcomes/characteristics of the various outputs and their valuation. An outcome is anything which is believed to affect welfare. Thus the most important outcome is the enhancement of health states measured by the change in the discounted quality adjusted life years enjoyed by patients as a result of treatments. Other outcomes include waiting time, quality of "hotel" services, patient satisfaction. Some care is needed in identifying independent outcomes. Thus patient satisfaction should not be included if it is believed to be driven entirely by quality of hotel services and the latter are included specifically.

For output j we identify the amount of each characteristic associated with it. The characteristics are measured in heterogeneous units. We denote the amount of characteristic m produced per unit of output j in period t as $q_{jm}(t)$. The price (marginal social value) associated with characteristic m in period t is $\pi_m(t)$. Then the total value of one unit of output j in period t is $\sum_m \pi_m(t) q_{jm}(t)$.

If $y_j(t)$ is the volume of outcome j, then the total social value of the product is $\sum_j \sum_m \pi_m(t) q_{jm}(t) y_j(t)$ and a base-weighted index of output growth is constructed as

(33) $$I_{Y(t)} = \frac{\sum_j \sum_m \pi_m(0) q_{jm}(t) y_j(t)}{\sum_j \sum_m \pi_m(0) q_{jm}(0) y_j(0)}$$

It should be noted that this takes account of quality change because it deals with changes in the amount of characteristic per unit of output as well as changes in the volume of each output.

If there is no up to date information on the amount of characteristic per procedure, then the index may take the form

(34) $$I_{Y^*(t)} = \frac{\sum_j \sum_m \pi_m(0) q_{jm}(0) y_j(t)}{\sum_j \sum_m \pi_m(0) q_{jm}(0) y_j(0)}$$

In this case it is simply an activity index with the weights being the base-period values put on those procedures through the process of valuing the characteristics of the procedures. If the NHS is reasonably efficient, so that these are not very different from social costs, such an index would not be very different from a cost-weighted activity index. If the characteristics are valued frequently, so that there is information on $\pi_m(t)$ for all periods, then it becomes possible to consider other index forms, such as the Fisher and Törnqvist indices. In practice it is unlikely that such data will be available with any frequency, at least in the near future.

Given this general specification, the question becomes how to measure the outputs y, their outcomes q and the value of the outcomes π.

5.2.2 Measurement of outputs

The measurement of outputs is the most straightforward though it is important to recognise that organisational divisions in the NHS mean that there will rarely be data on complete patient episodes taking place, as most do, across secondary and primary care. We will be investigating data availability more thoroughly in the next stage of the project but we note that whilst there appears to be is reasonably good data on

secondary care activities the primary care sector is much less well mapped. (See Appendix E.)

5.2.3 Measurement of characteristics per procedure

Our presumption is that the most important characteristic of most procedures is health gain, because the primary purpose of the health service is to provide treatment. Section 3.3 discusses the existing data and data requirements needed to ensure that this can be satisfactorily measured. The group will review the research evidence on key characteristics of health care which ought to enter the index of outcomes and will seek the advice of the Department of Health on available data.

5.2.4 Prices of characteristics

In section 3.5 we discussed a number of different ways in which prices can be attributed to the characteristics. The exercise is bound to be eclectic since different procedures will be sensible for different items. We note in particular that there is the residual option of valuing characteristics on the basis of cost of provision and that this should offer a method where none other is available.

5.3 Inputs and productivity

Denoting inputs by *z(t)* and their associated prices by *ω(t)*, a base-weighted index of aggregate input growth can be defined as:

$$I_{Z(t)} = \frac{\sum_n \omega_n(0) z_n(t)}{\sum_n \omega_n(0) z_n(0)} \tag{35}$$

Here there are fewer conceptual problems since inputs are sold on markets, hence prices are readily available. There may well be divergences between input prices and marginal social costs but it is unlikely that the project can address this issue within its time scale. Quality adjustments can partly be incorporated by suitable disaggregation and chain linking so that the use of say more skilled labour will be captured.

Capturing increases in quality due to the use of new and better drugs will be more difficult.

Productivity change is measured as the growth in the aggregate output index minus the growth in the aggregate input index. The results from the project will include not only a measure of productivity but also the contribution of growth in inputs, both aggregate and individually to explaining output growth.

5.4 Implementation

Given the methodological framework outlined above the next step is to consider how to implement this in practice. This will be a two stage procedure.

1. The research team will outline the data requirements and discuss their feasibility with the Department and other bodies such as ONS. This may lead to increased availability of data not readily in the public domain.
2. The research team then decide on which data series should be actually employed in attempting a first measure of outputs, inputs and productivity.

References

Aigner, D.J., Lovell, C.A.K. and Schmidt, P. (1977) 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics,* July, 6 (1): 21-37.

Almeida, C., Braveman, P., Gold, M.R., Szwarcwald, C.L., Ribeiro, J.M., Miglionico, A., Millar, J.S., Porto, S., Costa, N.R., Rubio, V.O., Segall, M., Starfield, B., Trasvessos, C., Uga, A., Valente, J. and Viacava, F. (2001) 'Methodological concerns and recommendations on policy consequences of the World Health Report 2000', *The Lancet,* 357 (9269): 1692-7.

Appleby, J. and Mulligan, J. (2000) 'How well is the NHS performing? A composite performance indicator based on public consultation', King's Fund, London.

Appleby, J. and Street, A. (2001) 'Health system goals: life, death and ... football', *Journal of Health Services Research,* 6 (4): 220-225.

Atkinson, T. (2004) 'Measurement of government output and productivity for the National Accounts', Atkinson Review: Interim Report, London, TSO.

Audit Commission (2003) 'Waiting for Elective Admissions', The Audit Commission.

Bates, J.M., Baines, D.L. and Whynes, D.K. (1996) 'Measuring the efficiency of prescribing by general practitioners', *Journal of the Operational Research Society,* 47 (12): 1443-51.

Bates, J.M., Baines, D.L. and Whynes, D.K. (1998) 'Assessing efficiency in general practice: an application of data envelopment analysis', *Health Services Management Research,* 11 (2): 103-8.

Battese, G.E. and Coelli, T.J. (1995) 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics,* 20: 325-332.

Berndt, E. (2003) 'Changes in the costs of treating mental health disorders: an overview of recent research findings', Tufts European Center Workshop, Preserving Access by Protecting Innovation, Taillores, France.

Berndt, E., Busch, S. and Frank, R.G. (2000/01) 'Interpreting changes in mental health expenditures: minding our Ps and Qs', NBER Research Summary.

Berndt, E.R., Bir, A., Busch, S.H., Frank, R.G. and Normand, S.T. (2002) 'The medical treatment of depression, 1991-1996: productive inefficiency, expected outcome variations, and price indexes', *Journal of Health Economics,* 21: 373-396.

Berndt, E.R., Busch, S.H. and Frank, R.G. (2001) 'Treatment price indexes for acute phase major depression', in D.M. Cutler and E.R. Berndt, eds.: *Medical Care Output and Productivity,* University of Chicago Press, Chicago.

Berndt, E.R., Cutler, D.M., Frank, R.G., Griliches, Z., Newhouse, J.P. and Triplett, J.E. (2000) 'Medical Care Prices and Output', in A.J. Culyer and J.P. Newhouse, eds.: *Handbook of Health Economics,* Volume 1, Elsevier Science B.V.

Besley, T., Hall, J. and Preston, I. (1999) 'The demand for private health insurance: do waiting lists matter?' *Journal of Public Economics,* 72: 155-181.

Bosch, J.L., van der Graaf, Y. and Hunink, M.G.M. (1999) 'Health-Related Quality of Life After Angioplasty and Stent Placement in Patients With Iliac Artery Occlusive Disease - Results of a Randomized Controlled Clinical Trial', *Circulation,* 24: 3155-3160.

Brilstra, E.H., Rinkel, G.J.E., van der Graaf, Y., Sluzewski, M., Groen, R.J., Lo, R.T.H. and Tulleken, C.A.F. (2004) 'Quality of Life after Treatment of Unruptured Intracranial Aneurysms by Neurosurgical Clipping or by Embolisation with coils', *Cerebrovascular Diseases,* 17 (1).

Buck, D. (2000) 'The efficiency of the Community Dental Service in England: a data envelopment analysis,' *Community Dentistry and Oral Epidemiology,* 28 (4): 274-80.

Burge, P., Devlin, N., Appleby, J., Rohr, C. and Grant, J. (2004) 'Do patients always prefer shorter waiting times? A discrete choice analysis of patients' stated preferences in the London Patient Choice Project (LPCP)', *Applied Health Economics and Health Policy,* 3 (1 (supp.)).

Canadian Institute for Health Information (2001a) 'Health Care in Canada 2001: A Second Annual Report', Canadian Institute for Health Information, Ottawa.

Canadian Institute for Health Information (2001b) 'Health Indicators 2001', Canadian Institute for Health Information, Ottawa.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982) 'Multilateral comparisons of output, input and productivity using superlative index numbers', *Economic Journal,* March, 92 (365): 73-86.

Cellini, R., Pignataro, G. and Rizzo, I. (2000) 'Competition and Efficiency in Health Care: An Analysis of the Italian Case', *International Tax and Public Finance,* 7 (4-5): 503-19.

Chambers, R., Fare, R. and Grosskopf, S. (1994) 'Efficiency, quantity indexes and productivity indexes: a synthesis', *Bulletin of Economic Research,* 46 (1): 1-21.

Chattapadhyay, S. and Ray, S.C. (1996) 'Technical, scale, and size efficiency in nursing home care: a nonparametric analysis of Connecticut homes', *Health Economics,* 5 (4): 363-73.

Chernew, M.E., Fendrick, M.A. and Hirth, R.A. (1997) 'Managed care and medical technology', *Health Affairs,* 16: 196-206.

Chilingerian, J.A. and Sherman, H.D. (1990) 'Managing physician efficiency and effectiveness in providing hospital services', *Health Services Management Research,* 3 (1): 3-15.

Chirikos, T.N. and Sear, A.M. (2000) 'Measuring hospital efficiency', *Health Services Research,* 34 (6): 1389-1408.

Cleary, P.D. and Edgman-Levitan, P. (1997) 'Health Care Quality', *Journal of American Medical Association,* 278 (19): 1608-1612.

Cockburn, I.M. and Anis, A.H. (2001) 'Hedonic Analysis of Arthritis Drugs', in D.M. Cutler and E.R. Berndt, eds.: *Medical Care Output and Productivity,* University of Chicago Press, Chicago.

Coelli, T.J. (1995) 'Measurement and sources of technical inefficiency in Australian electricity generation', Paper presented at the New England Conference on Efficiency and Productivity, University of New England, Armidale.

Coelli, T.J., Prasada Rao, D.S. and Battese, G.E. (1998) *An introduction to efficiency and productivity analysis*, Kluwer Academic Publishers, Boston.

Commission for Health Improvement (2003a) 'NHS performance ratings acute trusts, specialist trusts, ambulance trusts 2002/2003', Commission for Health Improvement, London.

Commission for Health Improvement (2003b) 'Rating the NHS: A question and answer guide to the Finsbury Rules, How CHI CGR scores will affect 2003 star ratings', Commission for Health Improvement, London.

Culyer, A.J., Lavers, R.J. and Williams, A. (1971) 'Social Indicators: Health', *Social Trends,* 2: 31-42.

Cutler, D.M. and Huckman, R.S. (2003) 'Technological development and medical productivity: the diffusion of angioplasty in New York state', *Journal of Health Economics,* 22: 187-217.

Cutler, D.M., McClellan, M. and Newhouse, J.P. (1999) 'The costs and benefits of intensive treatment for cardiovascular disease', in J.E. Triplett, ed.: *Measuring the prices of medical treatments,* The Brookings Institution, Washington, D.C.

Cutler, D.M., McClellan, M., Newhouse, J.P. and Remler, D. (2001) 'Pricing Heart Attack Treatments', in D.M. Cutler and E.R. Berndt, eds.: *Medical Care Output and Productivity,* University of Chicago Press, Chicago.

Cutler, T. (2002) 'Star or black hole?' *Community Care*: 40-41.

Dacosta, C.I. and Lapierre, S.D. (2003) 'Benchmarking as a tool for the improvement of health services' supply departments', *Health Services Management Research,* 16 (4): 211-23.

Deaton, A. and Muellbauer, J. (1980) *Economics and Consumer Behaviour*, Cambridge University Press, Cambridge.

Department of Health (2001) 'NHS Performance Ratings: Acute Trusts 2001/02', Department of Health, London.

Department of Health (2002a) 'NHS Performance Ratings and Indicators: Acute Trusts, Specialist Trusts, Ambulance Trusts, Mental Health Trusts, 2001/02', Department of Health, London.

Department of Health (2002b) 'Reforming NHS Financial Flows: introducing payment by results', Department of Health, London.

Department of Health (2004) 'Chief Executive's Report to the NHS', Department of Health, London.

Diewert, W.E. (1976) 'Exact and superlative index numbers', *Journal of Econometrics,* May, 4 (2): 115-145.

Dor, A. (1994) 'Non-minimum cost functions and the stochastic frontier: On applications to health care providers', *Journal of Health Economics,* 13: 329-34.

Eefting, F., Nathoe, H., van Dijk, D., Jansen, E., Lahpor, J., Stella, P., Suyker, W., Diephuis, J., Suryapranata, H., Ernst, S., Borst, C., Buskens, E., Grobbee, D. and de Jaegere, P. (2003) 'Randomized Comparison Between Stenting and Off-Pump Bypass Surgery in Patients Referred for Angioplasty', *Circulation,* 108 (23): 2870-2876.

Eurostat (2001) *Handbook on price and volume measures in National Accounts*.

Fanshel, S. and Bush, J.W. (1970) 'A health status index: its application to health service outcomes', *Operations Research,* 18: 1021-1066.

Farrell, M.J. (1957) 'The measurement of productive efficiency', *Journal of the Royal Statistical Society Series A,* 120: 253-90.

Frank, R., Berndt, E., Busch, S. and Lehman, A. (2003) 'Quality-constant price indexes for the ongoing treatment of schizophrenia: an exploratory study', *NBER working papers,* No. 10222.

Frank, R., Busch, S. and Berndt, E. (1998) 'Measuring prices and quantities of treatment for depression', *American Economic Review,* 88 (2): 106-111.

Friesner, D.L. (2003) 'An Empirical Examination of Cost-Adjusting in Outpatient Clinics', *Journal of Socio Economics,* 31 (6): 745-59.

Garcia, F., Marcuello, C., Serrano, D. and Urbina, O. (1999) 'Evaluation of efficiency in primary health care centres: an application of data envelopment analysis', *Financial Accountability and Management,* 15 (1): 67-83.

Garry, R., Fountain, J., Mason, S., Napp, V., Brown, J., Hawe, J., Clayton, R., Abbott, J., Phillips, G., Whittaker, M., Lilford, R. and Bridgman, S. (2004) 'The eVALuate study: two parellel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy', *British Medical Journal*: 1-7.

Gerdtham, U.G., Rehnberg, C. and Tambour, M. (1999) 'The Impact of Internal Markets on Health Care Efficiency: Evidence from Health Care Reforms in Sweden', *Applied Economics,* August, 31 (8): 935-45.

Giuffrida, A. and Gravelle, H. (2001) 'Measuring Performance in Primary Care: Econometric Analysis and DEA', *Applied Economics,* February, 33 (2): 163-75.

Gravelle, H., Jacobs, R., Jones, A.M. and Street, A. (2004) 'Comparing the efficiency of national health systems: a sensitivity analysis of the WHO approach', *Applied Health Economics and Health Policy,* forthcoming.

Gravelle, H. and Smith, P. (2001) 'Discounting for health effects in cost benefit and cost effectiveness analysis', *Health Economics,* 10: 587-600.

Gudex, C., Williams, A., Jourdan, M., Mason, R., Maynard, J., O'Flynn, R. and Rendall, M. (1990) 'Prioritising waiting lists', *Health Trends,* 22 (3): 103-108.

Harris, J., Ozgen, H. and Ozcan, Y. (2000) 'Do mergers enhance the performance of hospital efficiency?' *Journal of Operational Research Society,* 51 (7): 801-11.

Hicks, J.R. (1940) 'The valuation of social income', *Economica,* 7: 105-124.

HM Treasury (2004) 'Stability, security and opportunity for all: investing for Britain's long-term future', 2004 Spending Review, HM Treasury, London.

Hofmarcher, M.M., Paterson, I. and Riedel, M. (2002) 'Measuring Hospital Efficiency in Austria--A DEA Approach', *Health Care Management Science,* February, 5 (1): 7-14.

Hollingsworth, B. (2003) 'Non-parametric and Parametric Applications Measuring Efficiency in Health Care', *Health Care Management Science,* November, 6 (4): 203-18.

Hollingsworth, B., Dawson, P.J. and Maniadakis, N. (1999) 'Efficiency Measurement of Health Care: A Review of Non-Parametric Methods and Applications', *Health Care Management Science,* July, 2 (3): 161-72.

Hollingsworth, B. and Parkin, D. (2001) 'The efficiency of the delivery of neonatal care in the UK', *Journal of Public Health Medicine,* 23 (1): 47-50.

Hollingsworth, B. and Parkin, D. (2003) 'Efficiency and productivity change in the English National Health Service: can data envelopment analysis provide a robust and useful measure?' *Journal of Health Services Research and Policy,* 8 (4): 230-6.

Hollingsworth, B. and Wildman, J. (2003) 'The Efficiency of Health Production: Re-estimating the WHO Panel Data Using Parametric and Non-parametric Approaches to Provide Additional Information', *Health Economics,* June, 12 (6): 493-504.

Hulten, C.R. (2001) 'Total Factor Productivity: a short bibliography', in C.R. Hulten, E.R.Dean and M.J.Harper, eds.: *New Developments in Productivity Analysis,* NBER studies in income and wealth, Volume 63, University of Chicago Press, Chicago.

Jackson, J.L., Chamberlin, J. and Kroenke, K. (2001) 'Predictors of patient satisfaction', *Social Science and Medicine,* 52: 609-620.

Jacobs, R. (2001) 'Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis', *Health Care Management Science,* June, 4 (2): 103-15.

Jacobs, R., Smith, P.C. and Goddard, M. (2003) 'Measuring performance: An examination of composite performance indicators', A report for the Department of Health, Centre for Health Economics, University of York.

Jencks, S.F., Cuerdon, T., Burwen, D.R., Fleming, B., Houck, P.M., Kussmaul, A.E., Nilasena, D.S., Ordin, D.L. and Arday, D.R. (2000) 'Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels', *Journal of the American Medical Association,* 284 (13): 1670-6.

Jencks, S.F., Huff, E.D. and Cuerdon, T. (2003) 'Change in the quality of care delivered to Medicare beneficiaries, 1998-99 to 2000-2001', *Journal of the American Medical Association,* 289 (3): 305-12.

Jenkinson, C., Coulter, A., Bruster, S., Chandola, T. and Jones, P. (2003) 'Factors relating to patients' reports about hospital care for coronary heart disease in England', *Journal of Health Services Research & Policy,* 8 (2): 83-86.

Johnston, K. and Gerard, K. (2001) 'Assessing efficiency in the UK breast screening programme: does size of screening unit make a difference?' *Health Policy,* 56 (1): 21-32.

Jorgenson, D.W., Gollop, F.M. and Fraumeni, B. (1987) *Productivity and US Economic Growth*, Harvard University Press, Cambridge MA.

Jorgenson, D.W. and Griliches, Z. (1967) 'The explanation of productivity change', *Review of Economic Studies,* 34: 249-83.

Jorgenson, D.W. and Stiroh, K. (2000) 'Raising the Speed Limit: U.S. Economic Growth in the Information Age', *Brookings Papers on Economic Activity,* 1: 125-211.

Kind, P., Leese, B., Cameron, I. and Carpenter, J. (1993) 'Quantifying quality - measuring quality in the provision of health care', in M. Malek, P. Vacani, J. Rasquinha and P. Davey, eds.: *Managerial Issues in the Reformed NHS,* John Wiley & Sons Ltd, Chichester.

Kmietowicz, Z. (2003) 'Star rating system fails to reduce variation', *British Medical Journal,* 327 (7408): 184.

Kooreman, P. (1994) 'Nursing home care in The Netherlands: A nonparametric efficiency analysis', *Journal of Health Economics,* 13: 301-16.

Kumbhakar, S.C., Ghosh, S. and McGuckin, J.T. (1991) 'A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms', *Journal of business and economic statistics,* 9 (3): 279-286.

Kumbhakar, S.C. and Lovell, C.A.K. (2001) *Stochastic frontier analysis*, Cambridge University Press, Cambridge, New York.

Lancaster, K. (1971) *Consumer Demand: A New Approach.*, Columbia University Press, New York.

Lichtenberg, F.R. and Virabhak, S. (2002) 'Pharmaceutical-embodied technical progress, longevity, and quality of life: drugs as "equipment for your health"', *NBER working papers,* No. 9351.

Ling, D.Y.C., Busch, A.B., Frank, R.G. and Pakes, J. (2002) 'Treatment price indexes for bipolar disorder,' Poster presentation at National Institute of Mental Health, Economics of Mental Health workshop, Bethesda MD.

Linna, M. (1998) 'Measuring hospital cost efficiency with panel data models', *Health Economics,* 7: 415-427.

Lloyd, A. (in press) 'On patients undergoing carotid endarterectomy'.

Luoma, K., Järviö, M.L., Suoniemi, I. and Hjerppe, R.T. (1996) 'Financial Incentives and Productive Efficiency in Finnish Health Centres', *Health Economics,* Sept. Oct., 5 (5): 435-45.

Lynch, J.R. and Ozcan, Y.A. (1994) 'Hospital closure: an efficiency analysis', *Hospital and Health Services Administration,* 39 (2): 225-30.

Magnusson, J. (1996) 'Efficiency measurement and the operationalisation of hospital production', *Health Services Research,* 31 (1): 21-37.

Malmquist, S. (1953) 'Index numbers and indifference surfaces', *Trabajos de Estadistica,* 4: 209-42.

Manca, A., Sculphur, M.J., Ward, K. and Hilton, P. (2003) 'A cost-utility analysis of tension-free vaginal tape *versus* colposuspension for primary urodynamic stress incontinence', *The British Journal of Obstetrics and Gynaecology,* 110: 255-262.

McCallion, G., McKillop, D.G., Glass, J.C. and Kerr, C. (1999) 'Rationalising Northern Ireland hospital services towards larger providers: best-practice efficiency studies and current policy', *Public Money and Management,* 19 (2): 27-32.

Meeusen, W. and van Den Broeck, J. (1977) 'Efficiency estimation from Cobb-Douglas production functions with composed error', *International Economic Review,* 18 (2): 435-44.

Miller, N. (2002) 'Missing the target', *Community Care*: 36-38.

Mirmirani, S. and Li, H.C. (1995) 'Health Care Efficiency Measurement: An Application of Data Envelopment Analysis', *Rivista Internazionale di Scienze Economiche e Commerciali,* March, 42 (3): 217-29.

Nathoe, H.M., van Dijk, D., Jansen, E.W.L., Suyker, W.J.I., Diephius, J.C., van Boven, W., Brutel de la Riviere, A., Borst, C., Kalkman, C.J., Grobbee, D.E., Buskens, E. and de Jaegere, P.P.T. (2003) 'A Comparison of On-Pump and Off-Pump Coronary Bypass Surgery in Low-Risk Patients', *The New England Journal of Medicine,* 348 (5): 394-402.

Navarro, V. (2000) 'Assessment of the World Health Report', *The Lancet,* 356 (9241): 1598-1601.

Navarro, V. (2001) 'World Health Report 2000: response to Murray and Frenk', *The Lancet,* 357 (9269): 1701-02.

Navarro, V. (2002) 'The World Health Report 2000: can health care systems be compared using a single measure of performance?' *American Journal of Public Health,* 92 (1): 31-34.

Netten, A. and Curtis, L. (2003) *Unit costs of health and social care 2003*, Personal Social Services Research Unit, The University of Kent at Canterbury, Cornwallis Building, Canterbury.

Nord, E. (2002) 'Measures of goal attainment and performance: a brief, critical consumer guide', *Health Policy,* 59 (3): 183-91.

Nunamaker, T.R. (1985) 'Using data envelopment analysis to measure the efficiency of non-profit organisations', *Managerial and Decision Economics,* 6: 50-55.

Nyman, J.A., Bricker, D.L. and Link, D. (1990) 'Technical efficiency in nursing homes', *Medical Care,* 28 (6): 541-51.

O'Mahony, M. (1999) *Britain's Productivity Performance 1950-1996*, NIESR, London.

O'Mahony, M. and De Boer, W. (2002) 'Britain's relative productivity performance: has anything changed?' *National Institute Economic Review,* January: 72-82.

O'Mahony, M., Plant, H. and Stevens, P.A. (2003) 'Measuring the cost effectiveness of local authorities', Report to the ODPM.

O'Mahony, M. and Stevens, P. (2003) 'International comparisons of performance in public services: outcome based measures for education', mimeo, NIESR, London.

O'Mahony, M. and van Ark, B., eds. (2003) *EU Productivity and Competitiveness: An Industry Perspective. Can Europe Resume the Catching-up Process?* European Commission.

Orme, C. and Smith, P.C. (1996) 'The potential for endogeneity bias in data envelopment analysis', *Journal of the Operational Research Society,* 47: 73-83.

Ostendorf, M., Buskens, E., van Stel, H., Schrijvers, A., Marting, L., Dhert, W. and Verbout, A. (2003) 'Waiting for Total Hip Arthroplasty', *The Journal of Arthroplasty,* 19 (3): 302-309.

Ozcan, Y.A., Luke, R.D. and Haksever, C. (1992) 'Ownership and organizational performance: a comparison of technical efficiency across hospital types', *Medical Care,* 30 (9): 781-94.

Ozcan, Y.A., Watts, J., Harris, J. and Wogen, S.E. (1998) 'Provider experience and technical efficiency in the treatment of stroke patients: DEA approach', *Journal of the Operational Research Society,* 49 (6): 573-82.

Page, S. and Cramer, K. (2001) 'Maclean's Rankings of Health Care Indices in Canadian Communities, 2000: Comparisons and Statistical Contrivance', *Canadian Journal of Public Health,* 92 (4): 295-98.

Pauly, M.V. (1999) 'Medical care costs, benefits, and effects: conceptual issues for measuring price changes', in J.E. Triplett, ed.: *Measuring the prices of medical treatments,* Brookings Institution Press, Washington, D.C.

Pedraja-Chaparro, F., Salinas-Jiménez, J. and Smith, P. (1997) 'On the role of weight restrictions in data envelopment analysis', *Journal of Productivity Analysis,* 8 (2): 215-230.

Prinssen, M., Buskens, E. and Blankensteijn, J.D., on behalf of the DREAM trial participants (2004) 'Quality of Life after Endovascular and Open AAA Repair. Results of a Randomised Trial', *European Journal of Vascular and Endovascular Surgery,* 27: 121-127.

Prior, D. and Sola, M. (2000) 'Technical Efficiency and Economies of Diversification in Health Care', *Health Care Management Science,* 3 (4): 299-307.

Puig-Junoy, J. (1997) 'Measuring technical efficiency of output quality in intensive care units', *International Journal of Health Care Quality Assurance,* 10 (2/3): 117-120, 121-124.

Puig-Junoy, J. (1998) 'Technical efficiency in the clinical management of critically ill patients', *Health Economics,* 7 (3): 263-77.

Ramanathan, T.V., Chandra, K.S. and Thupeng, W.M. (2003) 'A Comparison of the Technical Efficiencies of Health Districts and Hospitals in Botswana', *Development Southern Africa,* June, 20 (2): 307-20.

Reifschneider, D. and Stevenson, R. (1991) 'Systematic departures from the frontier - a framework for the analysis of firm inefficiency', *International Economic Review,* 32 (3): 715-724.

Rosenman, R., Siddharthan, K. and Ahern, M. (1997) 'Output Efficiency of Health Maintenance Organizations in Florida', *Health Economics,* 6 (3): 295-302.

Rowan, K., Harrison, D., Brady, A. and Black, N. (2004) 'Hospitals' star ratings and clinical outcomes: ecological study', *British Medical Journal,* 328: 924-5.

Ryan, M., Odejar, M. and Napper, M. (2004) 'The Value of Reducing Waiting Time in the Provision of Health Care: A Review of the Evidence. A Report to the Department of Health'.

Scitovsky, A. (1964) 'An index of the cost of medical care - a proposed new approach', Proceedings of the conference on the Economics of Health and Medical Care, May 10-12, 1962, The University of Michigan.

Scitovsky, A. (1967) 'Changes in the costs of treatment of selected illnesses, 1951-65', *American Economic Review,* December: 1182-95.

Sculpher, M.J., Manca, A., Abbott, J., Fountain, J., Mason, S. and Garry, R. (2004) 'Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial', *British Medical Journal*: 1-6.

Shapiro, I., Shapiro, M.D. and Wilcox, D.W. (1999) 'Quality Improvement in Health Care: A Framework for Price and Output Measurement', *NBER working papers,* No. 6971.

Shapiro, I., Shapiro, M.D. and Wilcox, D.W. (2001) 'Measuring the Value of Cataract Surgery', in D.M. Cutler and E.R. Berndt, eds.: *Medical Care Output and Productivity,* University of Chicago Press, Chicago.

Simar, L. and Wilson, P.W. (2002) 'Estimation and inference in two-stage, semi-parametric models of production processes', Paper presented at Workshop on Quantitative Methods for the Measurement of Organizational Efficiency, Institute for Fiscal Studies, University College London.

Smith, P.C. (1997) 'Model misspecification in data envelopment analysis', *Annals of Operations Research,* 73: 233 - 252.

Smith, P.C. (2002) 'Developing composite indicators for assessing health system efficiency', in P.C. Smith, ed.: *Measuring up: Improving the performance of health systems in OECD countries,* OECD, Paris.

Snelling, I. (2003) 'Do star ratings really reflect hospital performance?' *Journal of Health Organization and Management,* 17: 210-223.

Solow, R. (1957) 'Technical change and the aggregate production function', *Review of Economics and Statistics,* 39: 312-320.

Stevens, P.A. (2001) 'The determinants of economic efficiency in English and Welsh universities', *NIESR Discussion Paper,* No. 185.

Stevens, P.A. (2004) 'Firms' expectations and employment behaviour: evidence from a panel of UK manufacturing firms', *Economics Letters,* forthcoming.

Tangelder, M.J.D., McDoennel, J., Van Bussbach, J.J., Buskens, E., Algra, A., Lawson, J.A. and Eikelboom, B.C. (1999) 'Quality of life after infrainguinal bypass grafting surgery', *Journal of Vascular Surgery,* 29 (5).

Thanassoulis, E. (1993) 'A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments', *Journal of the Operational Research Society,* 44 (11): 1129 - 1144.

Triplett, J.E. (1999) *Measuring the prices of medical treatments*, Brookings Institution Press, Washington D.C.

Triplett, J.E. (2001) 'What's Different about Health? Human Repair and Car Repair in National Accounts and in National Health Accounts', in D.M. Cutler and E.R. Berndt, eds.: *Medical Care Output and Productivity,* University of Chicago Press, Chicago.

Triplett, J.E. and Bosworth, B.P. (2002) *"Baumol's Disease" has been cured: IT and Multifactor productivity in US Services Industries*, Brookings Institution, Washington, D.C.

Triplett, J.E. and Bosworth, B.P. (2003) 'Productivity Measurement in Service Industries: 'Baumol's Disease' has been cured', *Federal Reserve Bank of New York Economic Policy Review,* 9 (3): 23-33.

Vallance-Owen, A. and Cubbin, S. (2002) 'Monitoring national clinical outcomes: a challenging programme', *British Journal of Health Care Management,* 8: 412-416.

Vitaliano, D.F. and Toren, M. (1994) 'Cost and efficiency in nursing homes: A stochastic frontier approach', *Journal of Health Economics,* 13: 281-300.

Wanless, D. (2002) 'Securing our Future Health: taking a long-term view', HM Treasury, London.

Williams, A. (1985) 'The economics of coronary artery bypass grafting', *British Medical Journal,* 291: 326-9.

Williams, A. (2001) 'Science or marketing at WHO?' *Health Economics,* 10: 93-100.

World Health Organization (2000) *The World Health Report 2000. Health systems: improving performance*, WHO, Geneva.

Yuen, P. and Towse, A. (2003) *NHS hospital episode statistics (HES) - increasing gap in measuring hospital activities*, Office of Health Economics, London.

Yule, G.U. (1934) 'On some points relating to vital statistics, more especially statistics of occupational mortality', *Journal of the Royal Statistical Society,* 97: 1-84.

Acknowledgements

The outputs and productivity project is funded by the Department of Health. The views expressed are not necessarily those of the DH. We are grateful to Mark Dusheiko for assistance with data and to Alan Williams, Alan Maynard, Peter Smith, Luigi Siciliani and Karen Bloor for helpful comments. We have also benefited greatly from advice from Jack Triplett who is a consultant to the project team.

Appendices

A Literature search strategy

Two electronic databases (HMIC and EconLit) were searched to identify potentially relevant papers.

HMIC

The HMIC (Health Management Information Consortium) database contains three separate databases:

1) DH-Data

This is the database of the Department of Health's Library & Information Services and contains in excess of 174,000 records relating to health and social care management information. Coverage includes official publications, journal articles and grey literature on: health service policy, management and administration, with an emphasis on the British National Health Service; the quality of health services including hospitals, nursing, primary care and public health; the planning, design, construction and maintenance of health service buildings; occupational health; control and regulation of medicines; medical equipment and supplies; and social care and personal social services. The majority of records are from 1983 onwards, although coverage of departmental materials dates back to 1919. Over a quarter of records have abstracts.

2) King's Fund Database

This database holds records of the material in the library of the King's Fund, an independent health charity working to develop and improve the management of health and social care services. This library is open to the public. Its database contains over 70,000 records (1979 to date), mostly with a UK focus, on health management and services, social care, service development, and NHS organisation and administration. Records include books, pamphlets, government reports, abstracts of journal articles and a wide range of 'grey' literature.

3) HELMIS

Health Management Information Service database,
Nuffield Institute for Health Information Resource Centre, Leeds University Library, University of Leeds, UK.

The Health Management Information Service (HELMIS) database is produced by the Nuffield Institute for Health Information Resource Centre of Leeds University Library. The database currently contains over 54,000 records (1984 to 1998) relating to community care and health systems management internationally but with particular focus on the United Kingdom, Europe (including Eastern Europe) and developing countries. Relevant items for inclusion are selected from UK and international journals, books, reports, official publications and "grey" literature. Records of journal articles include an abstract in English.

DH-Data and the King's Fund database include all the major health related academic journals.

HMIC search details:
Searched during April and May 2004.
Database date coverage: 1983 to May 2004
Search restricted to publication date 1989 onwards
Interface: WebSPIRS

An initial strategy was produced after discussion by the research team and, after consideration of the records retrieved, this was subsequently refined to ensure a closer focus on the research question. Both the strategies used are given below. The search was on title, key words and abstract.

Initial strategy

```
#1 quantity index
#2 growth accounting
#3 tornqvist index
```

#4 fisher index
#5 labour productivity
#6 residual productivity
#7 solow residual
#8 index number theory
#9 translog production
#10 hedonic
#11 medical productivity
#12 health productivity
#13 health care productivity
#14 healthcare productivity
#15 frontier analysis
#16 frontier analyses
#17 data envelopment analysis
#18 data envelopment analyses
#19 total factor productivity
#20 multifactor productivity
#21 multi factor productivity
#22 multi-factor productivity
#23 (multi-factor productivity) or (multi factor productivity) or (hedonic) or (translog production) or (index number theory) or (solow residual) or (residual productivity) or (labour productivity) or (fisher index) or (tornqvist index) or (growth accounting) or (quantity index) or (multifactor productivity) or (total factor productivity) or (data envelopment analyses) or (data envelopment analysis) or (frontier analyses) or (frontier analysis) or (healthcare productivity) or (health care productivity) or (health productivity) or (medical productivity)
#24 (medical service*) in ti,de
#25 (medical system*) in ti,de
#26 (public service*) in ti,de
#27 (public sector*) in ti,de
#28 ((health adj care) or healthcare) in ti,de
#29 (health service*) in ti,de
#30 (health system*) in ti,de
#31 (health insurance system*) in ti,de
#32 (primary care) in ti,de
#33 (family health service*) in ti,de
#34 (family practice*) in ti,de
#35 (general practice*) in ti,de
#36 nhs in ti,de
#37 (national health service*) in ti,de
#38 hospitals in ti,de
#39 ((medical service*) in ti,de) or (hospitals in ti,de) or ((national health service*) in ti,de) or (nhs in ti,de) or ((general practice*) in ti,de) or ((family practice*) in ti,de) or ((family health service*) in ti,de) or ((primary care) in ti,de) or ((health insurance system*) in ti,de) or ((health system*) in ti,de) or ((health service*) in ti,de) or (((health adj care) or healthcare) in ti,de) or ((public sector*) in ti,de) or ((public service*) in ti,de) or ((medical system*) in ti,de)
#40 (productivity or output or outputs of efficient or efficiency) in ti,de
#41 ((productivity or output or outputs of efficient or efficiency) in ti,de) and (((medical service*) in ti,de) or (hospitals in ti,de) or ((national health service*) in ti,de) or (nhs in ti,de) or ((general practice*) in ti,de) or ((family practice*) in ti,de) or ((family health service*) in ti,de) or ((primary care) in ti,de) or ((health insurance system*) in ti,de) or ((health system*) in ti,de) or ((health service*) in ti,de) or (((health adj care) or healthcare) in ti,de) or ((public sector*) in ti,de) or ((public service*) in ti,de) or ((medical system*) in ti,de))
#42 ((medical service*) in ti,de,ab) or (hospitals in ti,de,ab) or ((national health service*) in ti,de,ab) or (nhs in ti,de,ab) or ((general practice*) in ti,de,ab) or ((family practice*) in ti,de,ab) or ((family health service*) in ti,de,ab) or ((primary care) in ti,de,ab) or ((health insurance system*) in ti,de,ab) or ((health system*) in ti,de,ab) or ((health service*) in ti,de,ab) or (((health adj care) or healthcare) in ti,de,ab) or ((public sector*) in

ti,de,ab) or ((public service*) in ti,de,ab) or ((medical system*) in ti,de,ab)
#43 (productivity or output or outputs of efficient or efficiency) in ti,de,ab
#44 (measur* or estimat* or index or indexes or indices or indicator* or instrument* or calculat* or monitor* or gain* or chang* or increas* or improv* or grow*)in ti,de,ab
#45 ((productivity or output or outputs of efficient or efficiency) in ti,de,ab) and (((medical service*) in ti,de,ab) or (hospitals in ti,de,ab) or ((national health service*) in ti,de,ab) or (nhs in ti,de,ab) or ((general practice*) in ti,de,ab) or ((family practice*) in ti,de,ab) or ((family health service*) in ti,de,ab) or ((primary care) in ti,de,ab) or ((health insurance system*) in ti,de,ab) or ((health system*) in ti,de,ab) or ((health service*) in ti,de,ab) or (((health adj care) or healthcare) in ti,de,ab) or ((public sector*) in ti,de,ab) or ((public service*) in ti,de,ab) or ((medical system*) in ti,de,ab)) and ((measur* or estimat* or index or indexes or indices or indicator* or instrument* or calculat* or monitor* or gain* or chang* or increas* or improv* or grow*)in ti,de,ab)
#46 outcome* in ti
#47 ((medical service*) in ti) or (hospitals in ti) or ((national health service*) in ti) or (nhs in ti) or ((general practice*) in ti) or ((family practice*) in ti) or ((family health service*) in ti) or ((primary care) in ti) or ((health insurance system*) in ti) or ((health system*) in ti) or ((health service*) in ti) or (((health adj care) or healthcare) in ti) or ((public sector*) in ti) or ((public service*) in ti) or ((medical system*) in ti)
#48 (((medical service*) in ti) or (hospitals in ti) or ((national health service*) in ti) or (nhs in ti) or ((general practice*) in ti) or ((family practice*) in ti) or ((family health service*) in ti) or ((primary care) in ti) or ((health insurance system*) in ti) or ((health system*) in ti) or ((health service*) in ti) or (((health adj care) or healthcare) in ti) or ((public sector*) in ti) or ((public service*) in ti) or ((medical system*) in ti)) and (outcome* in ti)
#49 (technical change*) or (technological change*)
#50 (technical innovation*) or (technological innovation*)
#51 technology adj2 (chang*)
#52 technology adj (change or changes or changing or changed)
#53 ((technical innovation*) or (technological innovation*)) or (technology adj (change or changes or changing or changed)) or ((technical change*) or (technological change*)) or (technology adj2 (chang*))
#54 health sector account*
#55 (health or medical) near (price index)
#56 (health or medical) near (price indexes)
#57 (health or medical) near (price indices)
#58 ((measur* or estimate* or index or indexes or indices or indicator* or instrument* or calculat* or monitor* or gain* or chang* or increas* or improv* or grow*) in ti,de) near2 quality
#59 ((productivity or efficient or efficiency or output or outputs) in ti,de) near2 quality
#60 (((productivity or efficient or efficiency or output or outputs) in ti,de) near2 quality) or (((measur* or estimate* or index or indexes or indices or indicator* or instrument* or calculat* or monitor* or gain* or chang* or increas* or improv* or grow*) in ti,de) near2 quality)
#61 (inspection in ti) and ((services or council or borough or city) in ti)
#62 #60 not #61
#63 burden near2 disease
#64 cost* near2 disease*
#65 economic burden*
#66 (economic burden*) or (cost* near2 disease*) or (burden near2 disease)
#67 (survival rate*) near2 (measur* or estimate* or index or indexes or indices or indicator* or instrument* or calculat* on monitor* or gain* or chang* or increas* or improv* or grow*)
#68 physical capital

#69 capital near2 (hospital* or building* or equipment or infrastructure)
#70 (general practice* or family practice or family health service* or primary care) in ti,ab,de
#71 (input or inputs) and #70
#72 #23 or #41 or #45 or #48 or #53 or #54 or #55 or #56 or #62 or #57 or #66 or #67 or #68 or #69 or #71
#73 (#23 or #41 or #45 or #48 or #53 or #54 or #55 or #56 or #62 or #57 or #66 or #67 or #68 or #69 or #71) and ((PY:HMIC >= 1989) or (PY:HQ >= 1989)

Refined strategy

#1 quantity index
#2 growth accounting
#3 tornqvist index
#4 fisher index
#5 labour productivity
#6 residual productivity
#7 solow residual
#8 index number theory
#9 translog production
#10 hedonic
#11 medical productivity
#12 health productivity
#13 health care productivity
#14 healthcare productivity
#15 frontier analysis
#16 frontier analyses
#17 data envelopment analysis
#18 data envelopment analyses
#19 total factor productivity
#20 multifactor productivity
#21 multi factor productivity
#22 multi-factor productivity
#23 (multi-factor productivity) or (multi factor productivity) or (hedonic) or (translog production) or (index number theory) or (solow residual) or (residual productivity) or (labour productivity) or (fisher index) or (tornqvist index) or (growth accounting) or (quantity index) or (multifactor productivity) or (total factor productivity) or (data envelopment analyses) or (data envelopment analysis) or (frontier analyses) or (frontier analysis) or (healthcare productivity) or (health care productivity) or (health productivity) or (medical productivity)

#24((productivity OR output*) in ti,de) AND ((medical service*) in ti,de) or (hospitals in ti,de) or ((national health service*) in ti,de) or (nhs in ti,de) or ((general practice*) in ti,de) or ((family practice*) in ti,de) or ((family health service*) in ti,de) or ((primary care) in ti,de) or ((health insurance system*) in ti,de) or ((health system*) in ti,de) or ((health service*) in ti,de) or (((health adj care) or healthcare) in ti,de) or ((public sector*) in ti,de) or ((public service*) in ti,de) or ((medical system*) in ti,de)

#25 health sector account*

#26 (health or medical) near (price index)

#27 (health or medical) near (price indexes)

#28 (health or medical) near (price indices)

#29 #23 or #24 or #25 or #26 or #27 or #28 and ((PY:HMIC >= 1989) or (PY:HQ >= 1989)

EconLit

EconLit is a comprehensive, indexed bibliography with selected abstracts of the world's economic literature, produced by the American Economic Association. It includes coverage of over 400 major journals as well as articles in collective volumes (essays, proceedings, etc.), books, book reviews, dissertations, and working papers licensed from the Cambridge University Press Abstracts of Working Papers in Economics.

Over 99% of the articles are in English or include English summaries. Generally, all articles from all publications are indexed, including notes, communications, comments, replies, rejoinders, etc. Articles lacking author identification or without economic content are omitted.

EconLit search details:
Searched during April and May 2004.
Database date coverage: 1969 to May 2004
Search restricted to publication date 1989 onwards
Interface: WebSPIRS

An initial strategy was produced after discussion by the research team and, after consideration of the records retrieved, this was subsequently refined to ensure a closer focus on the research question. Both the strategies used are given below.

Initial strategy

```
#1 quantity index
#2 growth accounting
#3 tornqvist index
#4 fisher index
#5 labour productivity
#6 residual productivity
#7 solow residual
#8 index number theory
#9 translog production
#10 hedonic
#11 frontier analysis
#12 frontier analyses
#13 data envelopment analysis
#14 data envelopment analyses
#15 total factor productivity
```

#16 multifactor productivity
#17 multi-factor productivity
#18 multi factor productivity
#19 (hedonic) or (translog production) or (index number theory) or (solow residual) or (residual productivity) or (labour productivity) or (fisher index) or (tornqvist index) or (growth accounting) or (quantity index) or (multi factor productivity) or (multi-factor productivity) or (multifactor productivity) or (total factor productivity) or (data envelopment analyses) or (data envelopment analysis) or (frontier analyses) or (frontier analysis)
#20 medical service*
#21 medical system*
#22 public service*
#23 public sector*
#24 (health adj care) or healthcare
#25 health service*
#26 health system*
#27 health insurance system*
#28 primary care
#29 family health service*
#30 family practice*
#31 general practice*
#32 nhs
#33 national health service*
#34 hospitals
#35 ((health adj care) or healthcare) or (public sector*) or (public service*) or (medical system*) or (hospitals) or (national health service*) or (nhs) or (general practice*) or (medical service*) or (family practice*) or (family health service*) or (primary care) or (health insurance system*) or (health system*) or ((health service*) in ti,ab,de)
#36 (((health adj care) or healthcare) or (public sector*) or (public service*) or (medical system*) or (hospitals) or (national health service*) or (nhs) or (general practice*) or (medical service*) or (family practice*) or (family health service*) or (primary care) or (health insurance system*) or (health system*) or ((health service*) in ti,ab,de)) and ((hedonic) or (translog production) or (index number theory) or (solow residual) or (residual productivity) or (labour productivity) or (fisher index) or (tornqvist index) or (growth accounting) or (quantity index) or (multi factor productivity) or (multi-factor productivity) or (multifactor productivity) or (total factor productivity) or (data envelopment analyses) or (data envelopment analysis) or (frontier analyses) or (frontier analysis))
#37 ((health or medical) in ti) and #19
#38 (((health or medical) in ti) and #19) or ((((health adj care) or healthcare) or (public sector*) or (public service*) or (medical system*) or (hospitals) or (national health service*) or (nhs) or (general practice*) or (medical service*) or (family practice*) or (family health service*) or (primary care) or (health insurance system*) or (health system*) or ((health service*) in ti,ab,de)) and ((hedonic) or (translog production) or (index number theory) or (solow residual) or (residual productivity) or (labour productivity) or (fisher index) or (tornqvist index) or (growth accounting) or (quantity index) or (multi factor productivity) or (multi-factor productivity) or (multifactor productivity) or (total factor productivity) or (data envelopment analyses) or (data envelopment analysis) or (frontier analyses) or (frontier analysis)))
#39 (productivity or output or outputs or efficient or efficiency) near2 #35
#40 (productivity or output or outputs or efficient or efficiency) near2 (health or medical)
#41 ((productivity or output or outputs or efficient or efficiency) near2 #35) or ((productivity or output or outputs or efficient or efficiency) near2 (health or medical))
#42 outcome* and #35
#43 (technical innovation*) or (technological innovation*) or ((technology near2 (change or changes or changing or changed))in ti,ab,de)

```
#44 #43 and (health or medical)
#45 #43 and #35
#46 (#43 and #35) or (#43 and (health or medical))
#47 health sector account*
#48 (health or medical) near (price index or price indexes or price
indices)
#49 (measur* or estimat* or index or indices or indexes or indicator*
or instrument* or calculat* or monitor* or gain* or chang* or
increas* or improv* or grow*) near2 quality
#50 #49 and (health or medical or #35)
#51 burden near2 disease
#52 cost near2 disease*
#53 illness near2 burden
#54 (illness near2 burden) or (cost near2 disease*) or (burden near2
disease)
#55 survival rate*
#56 capital near2 hospital*
#57 (input or inputs) in ti,ab,de
#58 #57 near2 (health or medical or #35)
#59 (#38 or #41 or #42 or #46 or #47 or #48 or #50 or #54 or #55 or
#56 or #58) and (PY:ECON >= 1989)
```

Refined strategy

```
#1 ((health or medical) in ti) and ((hedonic) or (translog
production) or (index number theory) or (solow residual) or (residual
productivity) or (labour productivity) or (fisher index) or
(tornqvist index) or (growth accounting) or (quantity index) or
(multi factor productivity) or (multi-factor productivity) or
(multifactor productivity) or (total factor productivity) or (data
envelopment analyses) or (data envelopment analysis) or (frontier
analyses) or (frontier analysis))
#2 (((health adj care) or healthcare) or (public sector*) or (public
service*) or (medical system*) or (hospitals) or (national health
service*) or (nhs) or (general practice*) or (medical service*) or
(family practice*) or (family health service*) or (primary care) or
(health insurance system*) or (health system*) or ((health service*)
in ti,ab,de)) and ((hedonic) or (translog production) or (index
number theory) or (solow residual) or (residual productivity) or
(labour productivity) or (fisher index) or (tornqvist index) or
(growth accounting) or (quantity index) or (multi factor
productivity) or (multi-factor productivity) or (multifactor
productivity) or (total factor productivity) or (data envelopment
analyses) or (data envelopment analysis) or (frontier analyses) or
(frontier analysis))
#3 (productivity or output or outputs or efficient or efficiency)
near2 (health or medical)
#4 (productivity or output or outputs or efficient or efficiency)
near2 ((health adj care) or healthcare) or (public sector*) or
(public service*) or (medical system*) or (hospitals) or (national
health service*) or (nhs) or (general practice*) or (medical
service*) or (family practice*) or (family health service*) or
(primary care) or (health insurance system*) or (health system*) or
((health service*) in ti,ab,de)
#5 (productivity or output or outputs or efficient or efficiency)
near2 (health or medical)
#6 (#1 or #2 or #3 or #4 or 5) and (PY:ECON >= 1989)
```

B Frontier techniques: stochastic frontier and data envelopment analysis

Introduction to frontier methods

As stated in section 4.3, most studies of productivity assume that there is no inefficiency in the health service (or if there is, that it is unchanging). This assumption is unlikely to hold in practice. Studies have found inefficiency in many areas of government (e.g. McCallion *et al.* (1999) for hospitals, O'Mahony, Plant and Stevens (2003), for a wide range of local government services, and Stevens (2001, 2004) for higher education). Another assumption made is that there are no random influences on the measured levels of inputs and outputs. The NHS operates in an environment subject to considerable uncertainty; the precise time or place when certain influences occur – such as flu outbreaks, or spells of bad weather – is difficult or impossible to predict.

Measuring productivity change by the Laspeyres, Paasche or the superlative Fisher or Törnqvist indices requires quantity and price information as well as assumptions about the structure of technology and the behaviour of producers. Alternatively, change can be measured using a Malmquist (1953) productivity index, which does not require information or technological and behavioural assumptions. Unlike the other indices, however, the Malmquist index does require the estimation of a representation of the production technology (Coelli, Prasada Rao and Battese, 1998; Kumbhakar and Lovell, 2001). This production technology may be a production frontier, or its duals the cost or profit frontiers. The choice of method depends on the problem to be analysed. Once the production technology is estimated one can decompose TFP change into its two or three components: technical change, efficiency change, and possibly random influences. The two main empirical techniques that could be employed to do this are Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA). Although both techniques can be used to construct a Malmquist productivity index, as its name suggests, only stochastic frontier analysis can do so in a stochastic environment.

The Malmquist productivity index was introduced by Caves, Christensen and Diewert (1982) and is based on Malmquist's (1953) proposal of a quantity index for use in consumption analysis. Malmquist's (1953) index scaled consumption bundles up or down, in a radial fashion, to some arbitrarily selected indifference surface. As Chambers, Fare and Grosskopf (1994) show, Malmquist's scaling factor is in fact Shephard's (1953) input distance function, and Malmquist quantity indices for pair of consumption bundles can be constructed from ratios of corresponding pairs of input distance functions.

The fact that neither approach requires information on output prices is of particular use in the analysis of public sector bodies. In estimating the production technology, DEA and SFA obtain parameters which can be interpreted as shadow prices.

Stochastic frontier analysis

Stochastic frontier models date back to Aigner, Lovell and Schmidt (1977) and Meeusen and van Den Broeck (1977), who independently proposed a stochastic frontier production function with a two-part 'composed' error term. In the production context, where its use is most common, this error is composed of a standard random error term, representing measurement error and other random factors, and a one-sided random variable representing what Farrell (1957) called 'technical inefficiency', i.e. the distance of the observation from the production frontier. This notion of technical efficiency reflects the ability of a firm, hospital or unit to obtain maximal output from a given set of inputs. It is measured by the output of the firm relative to that which it could attain if it were 100% efficient, i.e. if it lay on the frontier itself, and is therefore bounded between zero and one. When one combines this with allocative efficiency (the ability of the firm etc to use the inputs in optimal proportions, given their respective prices) one has a measure of total *economic efficiency*.

Consider the single output case

$$Y_{it} = f\left(K_{it}, L_{it}, I_{it}; \boldsymbol{\beta}\right) \cdot E_{it} \cdot \exp\left(\upsilon_{it}\right) \tag{36}$$

where in this case Y_{it} is the output of unit i at time t, K = capital stock, L = labour, M =

materials, and the βs are parameters to be estimated, $0 < E \leq 1$ is an index of technical efficiency[2] and υ_{it} reflects the random character of the frontier, due to measurement error or other effects not captured by the model. Stochastic frontier analysis (SFA) commonly assumes that ln(E) (= ε) has a half-normal, truncated normal or gamma distribution and that υ is normal iid. In order to implement (36), we need to make an assumption regarding the functional form of the production frontier. For example, if we assume that production is Cobb-Douglas, taking logs and including a constant yields

$$y_{it} = \beta_{0t} + \beta_K k_{it} + \beta_L l_{it} + \beta_I i_{it} + \varepsilon_{it} + \upsilon_{it} \tag{37}$$

where lower case terms represent logs ($\varepsilon = \ln E$). In this case, TFP is equal to exp(β_{0jt} + υ_{it}). The change in output is

$$\Delta y_{it} = \beta_K \Delta k_{it} + \beta_L \Delta l_{it} + \beta_I \Delta i_{it} + \Delta\varepsilon_{it} + \Delta\upsilon_{it} \tag{38}$$

where the Δ operator denotes the log rate of change (e.g. $\Delta x_t = \ln x_t - \ln x_{t-1}$), and so

$$\Delta TFP = \Delta\varepsilon_{it} = \Delta y_{it} - \beta_K \Delta k_{it} + \beta_L \Delta l_{it} + \beta_I \Delta i_{it} + \Delta\upsilon_{it} \tag{39}$$

That is, the change in TFP is equal to the change in inefficiency. More generally, there may be three other sources of TFP change. If there are variable returns to scale there may also be a scale component to TFP change. If we have information on input prices, we can also calculate allocative inefficiency. However, perhaps the most important source of TFP change missing from (39) is the change in productivity due to technical change. In order to identify this, we must specify the source of technical change. Generally, technical change is defined as $T\dot{E}C = \partial \ln f(X, T; \beta) / \partial t$, where T represents the current level of technology and the β vector is increased to account for the respective parameters. For ease of exposition, we will assume that technical change is neutral with respect to the inputs (i.e. that $f(X, T; \beta) = A(T) \cdot g(X; \beta)$). If, for example, we assume that technology follows a deterministic linear time trend, (37) becomes

$$y_{it} = \beta_{0t} + \beta_K k_{it} + \beta_L l_{it} + \beta_I i_{it} + \beta_T t + \varepsilon_{it} + \upsilon_{it} \tag{40}$$

---

[2] $E = 1$ represents 100% efficiency, i.e. the unit produces on the frontier

and (39) becomes

$$\Delta TFP = \Delta\varepsilon_{it} = \Delta y_{it} - \beta_K \Delta k_{it} + \beta_L \Delta l_{it} + \beta_I \Delta i_{it} + \beta_T t + \Delta\upsilon_{it} \quad (41)$$

That is, technical change is constant over time (of course one could use a more general time trend, e.g. a quadratic or higher-order trend). A more satisfactory method would be to model the source of technical change, for example, following Kneller and Stevens (2001) we might specify that the current state of technical know-how is given by the stock of relevant knowledge. This could be measured by the number of patents relating to the area or (as in Kneller and Stevens, 2002) the stock of R&D capital. Thus the stochastic production frontier is

$$y_{it} = \beta_{0t} + \beta_K k_{it} + \beta_L l_{it} + \beta_I i_{it} + \beta_R R + \varepsilon_{it} + \upsilon_{it} \quad (42)$$

and TFP change is given by

$$\Delta TFP = \Delta\varepsilon_{it} = \Delta y_{it} - \beta_K \Delta k_{it} + \beta_L \Delta l_{it} + \beta_I \Delta i_{it} + \beta_R R + \Delta\upsilon_{it} . \quad (43)$$

Thus we can see that in order to utilise SFA in the measurement of health service inefficiency and productivity change, one must specify the source of technical change. Otherwise, we move from the growth accounting methodology where all productivity change is attributed to technical change, to a frontier-based one where no productivity change is attributed to changes in technology.

If there are a set of observable factors that we might expect to influence efficiency – for example, whether a hospital is attached to a university or whether staff are paid by fee-for-service or capitation – we can also estimate the significance or magnitude their effects. Using the method of Battese and Coelli (1995), the inefficiency effect is obtained by a truncation of the $N(\mu_{it}, \sigma^2)$-distribution, where

$$\mu_{it} = \boldsymbol{\delta}\boldsymbol{\Gamma}_{it} \quad (44)$$

where $\boldsymbol{\Gamma}_{it}$ is a vector of observable explanatory variables representing the efficiency determinants, and $\boldsymbol{\delta}$ is a vector of unknown scalar parameters to be estimated.

Although SFA is not able to account for multiple outputs in the primal (production frontier) form, it can when on uses the dual (cost frontier) approach. Estimation of productivity change becomes more complicated in the multiple output case. With more than one output, the Divisia index of productivity change becomes

$$\begin{aligned} \dot{TFP}_t &= \dot{\mathbf{Y}} - \dot{\mathbf{X}} \\ &= \sum_j R_j \dot{y}_j - \sum_n \varpi_n \dot{x}_n \end{aligned} \tag{45}$$

where $R_j = p_j y_j / R$ is the observed revenue share of output y_j, p_j is its price and $R = \Sigma_j p_j y_j$ is total revenue.

In the case of multiple outputs, we can estimate the cost dual to (36)

$$\ln \mathbf{XW} = \ln c(\mathbf{X}, \mathbf{T}; \boldsymbol{\beta}) + E + A \tag{46}$$

where **W** is the vector of input prices corresponding to **X**, (i.e. **XW** is total expenditure) and A is allocative inefficiency. In reality, it is difficult to decompose total economic inefficiency into its technical and allocative components, but relatively simple to obtain their sum (i.e. total economic efficiency). A common functional form for the cost function is the transcendental logarithmic, or translog.

$$\begin{aligned} \ln \mathbf{XW} = {} & \beta_0 + \sum_x \beta_{1x} \ln w_x + \sum_j \beta_{2j} \ln y_j + \sum_{x \neq z} {}_{2x}\ln w_x \cdot \ln w_z + \sum_j \sum_x {}_{2x}\ln w_x \cdot \ln Y_j \\ & + \tfrac{1}{2}\left[\beta_{3L} (\ln w_x)^2 + \beta_{4j} (\ln y_j)^2\right] + \eta + u \end{aligned} \tag{47}$$

where $\eta = \ln E + \ln A$. The translog is appealing because it provides a second order approximation to an arbitrary twice differentiable linearly homogeneous function (see Lau, 1974).

Data envelopment analysis

Data envelopment analysis (DEA) takes an entirely different approach to the estimation of the production technology. DEA estimates

(48) $$\mathbf{Y}_{it} = f\left(K_{it}, L_{it}, I_{it}; \boldsymbol{\beta}\right) \cdot E_{it}$$

by linear programming methods. By comparing (48) with (36), we can see two important differences between DEA and SFA. First, it can account for multiple outputs in the primal production frontier context. Second, DEA does not take into account a stochastic environment. DEA does not allow for allowances to be made for measurement error and other random effects. As we have suggested above, this is likely to particularly problematic in the health sector. DEA does not, however, have to impose a functional form on (48), although the value of this depends on what the underlying data generating mechanism is. Another potential advantage is its ability to deal with multiple outputs.

There are essentially three main problems with the application of DEA to the analysis of the health sector. A key problem with DEA is its heavy reliance on the accuracy of the sample data and the absence of random shocks; there is no allowance for stochastic errors. The second is whether a unit is efficient or merely unique. In some circumstances a unit may achieve a high efficiency score merely by being different (in its input or output mix) from other units. This is because, in effect, each unit chooses the criteria by which it wishes to be judged. Where the number of units under consideration is small this may lead to some units being deemed technically efficient simply because they are unusual. Related to this is the general issue of sensitivity. The relative efficiency score achieved by each unit can be sensitive to the number of inputs and outputs specified (Sexton, 1986; Nunamaker, 1985). The more input and output variables are included in the model, the higher will be the number of units with an efficiency score equal to unity (Nunamaker, 1985). In any application of DEA it is therefore important to test the sensitivity of the results to changes in input-output specification. However, there are no equivalents to the statistical tests used to obtain parsimonious econometric specifications that can be utilised in SFA.

The problems with sensitivity and uniqueness are closely related. DEA efficiency estimates are based on a comparison of the input-output levels of an individual unit with those of a very small subset of efficient peer units. They can therefore prove highly sensitive to data swings at the individual hospital level (Smith, 1997). If a hospital has no peers because it produces some unique output, it will end up on the

efficiency frontier, even though its classification as fully efficient may be questionable.

In most applications of DEA, the weights are allowed to vary across organisations. For each organisation, DEA computes all possible sets of weights and chooses those weights that assign the highest efficiency score. This flexibility significantly undermines conclusions that can be drawn about relative efficiency (Pedraja-Chaparro, Salinas-Jiménez and Smith, 1997). One way to overcome this is to impose weights, but such weights will necessarily be arbitrary or at least depend upon variables that would be included in a cost function, hence removing the flexibility advantage DEA has over SFA.

Another ongoing debate among proponents of DEA is the matter of how to control for the fact that organisations operate in diverse environmental contexts (Buck, 2000; Ozcan, Luke and Haksever, 1992). A common way of dealing with this in the literature has been to analyse efficiency scores in a second-stage econometric analysis (Gerdtham, Rehnberg and Tambour, 1999; Kooreman, 1994; Luoma *et al.*, 1996). This is not an appropriate for SFA type models, as Kumbhakar, Ghosh and McGuckin (1991) and Reifschneider and Stevenson (1991) have noted. In the first stage, when the production or cost relationship is estimated, the efficiency terms are assumed to be independently and *identically* distributed, but in the second stage they are assumed to be a function of these firm-specific factors, implying that they are not identically distributed[3]. There are also significant concerns with using this approach in DEA-based analyses, in that there is little theoretical justification for the choice of variables at each stage (Dor, 1994) and more notably the parameter estimates and the standard errors from these second-stage regressions are biased since the efficiency scores from DEA are serially correlated (Simar and Wilson, 2002).

One of the advantages, but also a disadvantage, of DEA is that it is perceived to be simple to implement and interpret. In a survey of a group of providers and purchasers in the NHS, when presented data on DEA and how to use it, 80 percent of those surveyed gave high scores for its potential usefulness compared with 9-45 percent for

[3] unless all the coefficients of the factors are simultaneously equal to zero

existing methods of efficiency (Hollingsworth and Parkin, 2003).

The application of frontier techniques to the health sector

There has been a burgeoning of literature applying these techniques to various healthcare settings (Hollingsworth, 2003; Hollingsworth, Dawson and Maniadakis, 1999). They have been used to examine the efficiency of health care systems (Hollingsworth and Wildman, 2003), individual hospitals (McCallion *et al.*, 1999), primary health care centres (Garcia *et al.*, 1999), general practices (Bates, Baines and Whynes, 1998), hospital wards (Hofmarcher, Paterson and Riedel, 2002), intensive care units (Puig-Junoy, 1998), neonatal units (Hollingsworth and Parkin, 2001), breast screening units (Johnston and Gerard, 2001), nursing homes (Chattapadhyay and Ray, 1996; Nyman, Bricker and Link, 1990; Vitaliano and Toren, 1994), and individual physicians (Chilingerian and Sherman, 1990). The methods have been applied in a number of countries including developing country settings (Ramanathan, Chandra and Thupeng, 2003) and have been used to study a range of issues such as economies of scale and scope (Dacosta and Lapierre, 2003; Magnusson, 1996), the impact of certain healthcare reforms (Gerdtham *et al.*, 1999; Mirmirani and Li, 1995), ownership types (Rosenman, Siddharthan and Ahern, 1997), competition (Cellini, Pignataro and Rizzo, 2000), mergers (Harris, Ozgen and Ozcan, 2000), hospital closures (Lynch and Ozcan, 1994), technology use (Puig-Junoy, 1997), diversification of hospital output (Prior and Sola, 2000), supply-chain management (Dacosta and Lapierre, 2003), and issues such as the change to the GP contract, and GP prescribing patterns (Bates, Baines and Whynes, 1996).

DEA and SFA are concerned with measuring the relative efficiency of organisational units and hence do not assume that all firms are technically efficient. The methods are most often applied to data on a sample of organisations at a particular point in time, but can also be applied to panel data to measure both technical change and efficiency change.

However, in an empirical analysis comparing these two methods to construct TFP indices (Coelli, 1995), the results diverged quite dramatically suggesting some cause

for concern in applying these methods without careful consideration. This is often the case when comparing DEA and SFA even for fairly simple production techniques and when the underlying models are equivalent (Chirikos and Sear, 2000; Giuffrida and Gravelle, 2001; Jacobs, 2001; Linna, 1998; Thanassoulis, 1993). These divergences are the result of the assumptions underpinning the DEA and SFA techniques in terms of how the methods establish the location and shape of the frontier.

Conclusions

Frontier methods have a number of characteristics that are appealing to the study of productivity. First they provide a framework within which productivity change can be decomposed into its components.

The fact that neither approach requires information on output prices is of particular use in the analysis of public sector bodies. In estimating the production technology, DEA and SFA obtain parameters which can be interpreted as shadow prices

Both methods are vulnerable to measurement and misspecification error with the dangers faced by all empirical analysis, such as omitting important variables (previously these have notably been quality (Friesner, 2003), casemix (Hollingsworth and Parkin, 2001), or severity (Ozcan *et al.*, 1998)), the inclusion of irrelevant variables (DEA), endogeneity bias, the adoption of an inappropriate technology (in SFA), or the imposition of an inappropriate variable returns to scale (VRS) assumption (in DEA) (Orme and Smith, 1996; Smith, 1997).

C Use of composite indicators in health care: international examples

The development of composite indicators may be relevant to the analysis of productivity in healthcare because composite indicators are essentially a weighted aggregated measure of healthcare outputs (performance indicators) where the composite C may take the following form:

$$C = \sum_{i=1}^{n} w_i x_i \tag{49}$$

where n is the number of indicators, w_i is the weight attached to indicator i, and x_i the score on indicator i.

Several of the steps in the construction of composite indicators may therefore be relevant to the productivity methodology, namely:

1. choosing the entities to be assessed
2. choosing the organisational objectives to be encompassed
3. choosing the performance indicators to be included (the x's)
4. transforming measured performance on individual indicators
5. combining the individual measures
6. specifying an appropriate set of weights

The following section reviews some of the literature on the development of composite indicators in healthcare and the problems with the composite indicator methodology.

United States Medicare

In 2000 and 2003 Jencks *et al.* produced a series of 22 quality indicators of the care delivered to Medicare beneficiaries (primarily in fee-for-service) and constructed state-level composite indices from these indicators. They examined these state-level indicators for the periods of 1998/99 (baseline) and 2000/01 (follow-up) and examined changes in performance across the range of indicators. The quality indicators were abstracted from state-wide random samples of medical records for

inpatient fee-for-service care (16 indicators) and from Medicare beneficiary surveys or Medicare claims for outpatient care (6 indicators).

Indicators were chosen for 6 clinical areas, namely: acute myocardial infarction (6 indicators), heart failure (2 indicators), stroke (3 indicators), pneumonia (7 indicators), breast cancer (1 indicator), and diabetes (3 indicators). The choice of indicators tended to over-represent inpatient and preventive services and under-represent ambulatory care and interventional procedures. The indicators were also not risk-adjusted and hence focused on process measures rather than outcomes. Clinical topics were selected according to five criteria:

- The disease is a major source of morbidity or mortality;
- Certain processes of care are known to improve outcomes;
- Measurement of these processes is feasible;
- There is substantial scope for improvement in performance;
- Managerial intervention can potentially improve performance.

In the first study (Jencks *et al.*, 2000), each of the 52 states is ranked on each of the measures, thus the percentage score is transformed to an ordinal scale ranging from 1 to 52. A composite performance measure was produced by computing each state's average rank.

In the follow-up study (Jencks, Huff and Cuerdon, 2003), absolute improvement was also calculated on each indicator (defined as the percentage improvement from baseline to follow-up) and relative improvement, or the reduction in error rate, (defined as the absolute improvement divided by the difference between baseline and perfect performance (100%)).

To summarize the overall changes in performance at the state level, they calculated the median amount of absolute and relative improvement across the set of indicators in the state. They also calculated the rank of each state on each quality indicator based on the 2000/01 performance and the rank on each quality indicator based on relative improvement. They then calculated the average rank for each state across the 22

indicators and league-tabled them according to their average rank based on 2000/01 performance as well as relative improvement. They found that a state's average rank on the 22 indicators was highly stable over time with a correlation of 0.93 between the two periods.

The better performing states appeared to be concentrated geographically in the northern and less populated regions (for both periods) while the geographic patterns of relative improvement by state were more patchy. While the report showed overall improvement across 20 of the 22 indicators and a median relative improvement of 12.8 percent (in the median state), the cross-sectional data used could not provide information about the source of the quality improvement and the role of quality improvement efforts in that.

There were some concerns about the choice of indicators (process measures) which the authors acknowledged as well as the data reliability of some of the indicators (Smith, 2002). Furthermore, since all indicators were given equal weight, the use of eight indicators for AMI would give that clinical area more of a contribution in the composite than say breast cancer for which there is only one indicator.

The use of the league table ranking as the basis for the composite also implicitly assumes that identical differences in ranking are equally important, regardless of where in the league table they occur. The incentive is therefore for states to concentrate on activities where they can more readily secure a movement up the league table, rather than those that offer the most potential health gain.

Canadian regional health care

Macleans magazine (MM) is a major mass circulation Canadian magazine that publishes an annual "Health Report" in which they rank Canadian regions according to their healthcare. This is done on the basis of data published by the Canadian Institute for Health Information on a series of annual reports as well as a series of health indicators for the 63 largest regions (covering 90 percent of the population) (Canadian Institute for Health Information, 2001a; Canadian Institute for Health Information, 2001b).

The third *Macleans* report in 2001 used 15 health care performance indicators grouped into six categories:

1. *Outcomes*: 1) Life expectancy at birth, 2) Heart attack survival;
2. *Prenatal care*: 1) Proportion low birthweight babies under 2500g, 2) Percentage of babies born by Caesarean section, 3) Percentage of vaginal births after Caesarean section;
3. *Community health*: 1) Hip fractures, 2) Pneumonia and flu hospitalisation of persons over 64;
4. *Elderly services*: 1) Hip replacements, 2) Knee replacements;
5. *Efficiency*: 1) Possible outpatients – hospitalisations for conditions not requiring admission, 2) Early discharge – variation from expected length of stay, 3) Preventable admissions - hospitalisations for conditions considered preventable by appropriate ambulatory care; and
6. *Resources*: 1) Physicians per 100,000; 2) Specialists per 100,000; 3) Local services – percentage of hospitalisations generated by local residents.

The MM report evaluated 54 regions with populations over 125,000, classified as either: communities with medical schools, other major communities, or largely rural communities.

MM rescaled each of the 15 indicators to have a mean of 80 and a standard deviation of 10 (with a higher score implying better performance). Where there was missing data, scores were inferred from performance on non-missing data. Within each of the six categories, the scores on the performance indicators were combined using weights 'based on expert opinion'. The six categories were then combined using the following weights: outcomes 0.2; prenatal care 0.2; community health 0.2; elderly services 0.1; efficiencies 0.2; resources 0.1. This sequential approach to assigning weights, first to performance indicators within categories and then to categories, allows a more careful treatment of priorities, albeit that the weighting scheme is very rudimentary and the preferences of the 'experts' may not necessarily reflect those of the public.

The composite performance scores ranged from 89.5 in North/West Vancouver, British Columbia to 73.4 in North Bay/Huntsville, Ontario in 2001.

MM acknowledge that there is very little variation in some of the indicators, however they discuss these variations as if clear differences exist which reflect real differences in health services. MM includes no psychiatric performance indicators which is a major aspect of health care services (Page and Cramer, 2001). Moreover the inclusion of efficiency indices leads efficiency to be treated not as the extent to which objectives are secured in relation to expenditure, but as simply another objective that contributes to the concept of performance (Smith, 2002).

British Health Authorities

In 2000, the UK television broadcaster Channel 4 commissioned researchers at the King's Fund to explore the public's relative preferences or health care priorities (Appleby and Mulligan, 2000). They produced a ranked list of English and Welsh health authorities and Scottish health boards according to a composite indicator based on selected aspects of performance, designed to reflect the relative weight attached by the public to these measures of NHS performance. Researchers were limited on the number of indicators with which they could feasibly survey the public and thus restricted their choice to six indicators chosen from readily available data produced by the NHS Executive (the High Level Performance Indicator set from the Performance Assessment Framework (PAF)):

1. Number of deaths from cancer (per 100,000)
2. Number of deaths from heart disease (per 100,000)
3. Number of people on hospital waiting lists (per 1,000)
4. Percentage of people on waiting lists for more than 12 months
5. Number of hip operations (per 100,000)
6. Number of deaths from 'avoidable' diseases (tuberculosis, asthma etc. for which there are effective clinical interventions that would prevent death) (per 100,000)

The focus of the study was to attach weights to each indicator based on public preferences. A polling organisation, MORI, surveyed 2000 people across England, Scotland and Wales to obtain their preferences. Three methods were used for eliciting preferences, namely ranking from most to least desired indicator, budget-pie, where

respondents were asked to allocate a 'budget' of 60 chips between the six performance indicators and conjoint analysis, a technique asking respondents to choose between different mixes of options. Statistical analysis was then used to extract the average value placed on each indicator by the respondents. All three methods produced very similar overall rankings of health authorities and health boards based on the final composite indicator.

The weights for each indicator were taken from the budget-pie method and are shown in Table C.1 with the weighting for hip operations taking a negative value indicating the "more is better" nature of the value. The weights were then multiplied by the actual performance values for each health authority and health board and then summed to produce a composite score, weighted by public preferences.

Table C.1 Weights for six performance indicators based on 'budget-pie' survey

Indicator	Number of chips	Weight
Reducing deaths from cancer	16	1.00
Reducing deaths from heart disease	12	0.75
Reducing total number of people on hospital waiting lists	10	0.63
Reducing number of people waiting more than 12 months	9	0.56
Reducing deaths from 'avoidable' diseases	5	0.50
Increasing number of hip operations	8	-0.31

Source: Appleby & Mulligan (2000)

The researchers were concerned that some of the raw performance indicators had skewed distributions and were not all measured on the same scale. The transformed data were standardised to a z score. The rankings generated from these transformed and standardised indicators differed from the original ranks with the average change in ranking being 14 places and a 0.81 rank correlation. The researchers then explored the extent to which socio-economic factors explained the variation in performance by controlling for socio-economic characteristics of the population. They controlled for deprivation using the Under-Privileged Area (UPA) score which explained 43 percent of the variation.

Again some concerns may be raised about the choice of indicators and their coverage

as well as the sensitivity to the rankings of health authorities from the transformation of the data.

Health authorities have now been superseded by strategic health authorities and the composite indicator system now in place is the star rating system for NHS provider organisations.

The World Health Report 2000

The composite index of health system performance produced for 191 countries by the World Health Organisation (WHO) in the *World Health Report* 2000 has been the subject of much debate (World Health Organization, 2000). The indicator was based on a weighted sum of attainment across 3 broad areas: health, financial fairness and responsiveness. Five dimensions were captured in total as the health and responsiveness indicators considered both the overall level of attainment and their distribution:

- Overall health outcomes
- Inequality in health
- Fairness of financing
- Overall health system responsiveness
- Inequality in health system responsiveness

The index has been widely discussed in the literature (Appleby and Street, 2001; Nord, 2002; Smith, 2002; Williams, 2001) and is described only briefly here.

The first dimension, average population health, was measured in terms of disability-adjusted life expectancy (DALE) that involves the calculation of severity weights for illnesses. The latter used the responses from an international panel of health personnel. The second dimension, equity of health, was measured primarily in terms of equality in survival for the first 5 years of life. The third dimension, fairness in financing, was measured by creating an index ranging from 0 to 1, defined as the ratio between total expenditure on health (tax plus out-of-pocket) and total non-food expenditure. The fourth dimension, responsiveness, was meant to capture how well the health care system responded to basic non-health expectations in the population

and was expressed as an index covering seven aspects of responsiveness (dignity and autonomy of patients, freedom of choice of provider, quality of amenities, prompt attention, confidentiality, access to social support networks). Each dimension was scored by around 2000 "key informants" from 35 countries who answered questions about their own country and were then asked to give a score for the aspect as a whole. The seven aspects were then ranked in order of importance by 1000 people and weights assigned based on the rankings. Mean scores on each aspect were multiplied by weights and summed to give an overall responsiveness score. For the other countries that were not asked directly to answer the responsiveness questions, scores were calculated by using the means of variables that had been shown to correlate strongly in the 35 countries. The fifth dimension equity in responsiveness was calculated by asking informants to make judgements about the subgroups they thought were treated with less responsiveness than others. Scores were assigned to sub-groups based on the number of times the country informants mentioned them, multiplied by the share of that group in the population. The products were summed and transformed to give an overall score. Scores for countries other than the 35 questioned were estimated in a similar way to the levels of responsiveness.

Finally, the scores on each of the 5 dimensions were transformed to a 0-100 scale using the formula in the table below and summed using the weights shown in the table, reflecting their relative importance, based on views of about 1000 people from 123 countries, half of whom were WHO staff.

Table C.2 Weights and transformations for the five objectives, WHO (2000)

Objective	Weight	Transformation
H: Overall health outcomes	0.250	(H-20)/(80-20)*100
HI: Inequality in health	0.250	(I-HI)*100
FF: Fairness of financing	0.250	FF*100
R: Overall health system responsiveness	0.125	(R/10)*100
RI: Inequality in health system responsiveness	0.125	(I-RI)*100

Source: World Health Organisation (2000)

A second index was also created using the WHO composite measure of attainment to

estimate overall health system efficiency. In this second stage of econometric modelling (using stochastic frontier analysis) an attempt was made to capture relative performance by looking at the difference between what was actually achieved in each country (attainment), compared with the maximum they could be expected to achieve given their resources. The latter was measured by health system expenditure and exogenous influences (human capital captured by years of education). A minimum level was also set and the index of relative performance was calculated by dividing the difference between actual attainment and minimum level by the difference between the maximum and minimum levels.

The debate about the appropriateness of the WHO rankings and methodology has been widespread (Navarro, 2000; Navarro, 2001; Navarro, 2002; Williams, 2001). Several criticisms relate to the lack of transparency about the methods used, with many commentators noting that it is not easy to find out from the WHO documentation how data were collected and transformed or the characteristics of respondents whose views were taken into account for the measures of responsiveness and the weights (Almeida *et al.*, 2001).

The major methodological concerns about the creation of the composite measure have revolved around the nature of the underlying data; the transformation of the data; and the weighting system; and how the weights were elicited (whether respondents' valuations of achievement against objectives truly represent marginal valuations). From an econometric perspective there are also contentious issues around the choice of functional form, the choice of error structure, the choice of covariates and the treatment of exogenous factors. The stochastic frontier method can be highly sensitive to model specification (Gravelle *et al.*, 2004). A huge number of technical and analytical judgements were therefore made in the WHO rankings.

The UK star rating system for NHS providers

In September 2001, the first set of performance star ratings were published by the Department of Health for acute NHS Trusts for 2000/01 (Department of Health, 2001). The star ratings are a composite index score given to each NHS organisation which are supposed to provide an overall assessment of performance across a number

of indicators. In July 2002, the second set of star ratings were published by the Department of Health, now covering acute Trusts, specialist Trusts, ambulance Trusts and indicative ratings for mental health Trusts for 2001/02 (Department of Health, 2002a). Primary Care Trusts (PCTs) received a separate publication, describing their performance against a range of suitable indicators, but not a rating. In July 2003, the most recent set of star ratings were published for 2002/03, covering again all types of NHS Trusts and PCTs. In this third round, the Commission for Health Improvement (CHI), the independent regulator of NHS performance, took over responsibility for performance ratings and indicators from the Department of Health (Commission for Health Improvement, 2003a).

The methodology for the three years of star ratings has remained relatively constant, although some important changes have been made to the individual indicators covered. There were broadly four areas of indicators in 2000/01: Clinical effectiveness and outcomes; Efficiency; Patient/carer experience; and Capacity & capability. In the last two years of star ratings the key areas have been: Key government targets; Clinical focus; Patient focus; Capacity & capability; and the CHI reviews.

The NHS Performance Ratings system places NHS Trusts in England into one of four categories, from three stars (the highest levels of performance) to zero stars (the poorest performance against indicators).

The key government targets are the most significant factors in determining overall performance ratings. The broader range of indicators make up a 'balanced scorecard' to refine the judgement on ratings and are combined in a complex 6-step process to produce the star ratings. CHI reviews of Trust clinical governance arrangements also play an important role in determining star ratings since three star Trusts need to perform well on all key targets as well as the CHI clinical review.

Performance against key targets is assessed in terms of whether the target has been achieved, whether there has been some degree of underachievement or whether the target was significantly underachieved.

The methodology broadly entails transforming the underlying key targets and performance indicators from continuous variables into categorical variables of either 3 or 5 categories. The performance indicators in the Patient, Clinical, and Capacity & Capability focus areas are categorised into one of five performance bands, with 5 points awarded for the best performance and 1 for the worst. The thresholds for deciding the cut-offs are not necessarily the same for each variable. The default position is to split performance into 5 bands by percentile.

Individual band scores are combined to produce an overall score per area. All indicators are equally weighted within their scorecard area in such a way as to ensure that despite differing numbers of indicators, each scorecard area carries the same weight.

The clinical indicators (in the Clinical focus area) are published with confidence intervals which means that performance is split into three bands dependent on whether the organisation's confidence interval overlaps the England average for the indicator (scoring 3) is significantly below average (scoring 1) or significantly above average (scoring 5).

The CHI clinical review is used to determine poorly performing (zero star) and high performing (three star) NHS organisations. A CHI clinical governance review assesses the trust across seven components of performance: risk management, clinical audit, research and education, patient involvement, information management, staff involvement, and education, training and development. Each component is scored from I to IV.

The role of the CHI clinical governance review has evolved over time. In 2001, the assessment for Star Ratings required only that the organisation had not received a critical review. The 2003 NHS Star Ratings for acute, specialist and mental health trusts were adjusted in accordance with the 'Finsbury rules' (Commission for Health Improvement, 2003b). In essence, these involve zero-rating any organisation that is evaluated as having achieved only the lowest standard of performance (level 'I') in five or more out of the seven areas of clinical governance assessed, apparently irrespective of the organisation's performance on key targets or the scorecard. Three

stars are awarded only to organisations that have achieved key targets, a balanced scorecard, at least three 'III's and no 'I's in the CHI review (Commission for Health Improvement, 2003b).

A six-step decision algorithm is then imposed whereby a sequential set of decisions on achievement on the various key variables determines the ultimate star rating outcome. The most important driving factors for obtaining the highest rating are the key targets and the CHI review which enter at several of the decision steps for Trusts to achieve. These therefore implicitly receive the highest weighting in the determination of the star rating.

There are many decisions in the process which may impact on whether Trusts are accorded a particular rating, in particular, the choice of performance indicators, the thresholds for transforming the performance indicators into categorical variables, the decision rules which are applied and the resultant implicit weighting given to each of the indicators. The star ratings have been shown to be sensitive to these methodological choices (Jacobs *et al.*, 2003).

The ratings are intended to be 'not primarily a commentary on the quality of clinical care', but rather to assess the 'overall patient experience' (Department of Health, 2001). However, both the construction and the impact of Star Ratings have been questioned (Cutler, 2002; Kmietowicz, 2003; Miller, 2002; Rowan *et al.*, 2004; Snelling, 2003).

D Price indices for health services: US developments

The problem of calculating price indices for health care is closely related to the problem of calculating output indices. If health services are produced in the private sector, then if we have price indices we can apply them to indices of consumers' expenditure to obtain output volume indices for calculating productivity growth. However if prices are incorrectly measured, due to changes in quality for example, output will also be incorrectly measured. Although there are no price indices for the NHS outputs, many of the issues confronted by writers in the US are relevant to the measurement of NHS output. As in the UK, there is concern over the increase in expenditure on health care, with many official statistics reporting price increases for many types of care.

What emerges from the recent literature is that it is important to focus on the direct medical costs of treating an episode of an illness, rather than changes in the prices of the inputs used in treatment, and that there needs to be some adjustment for quality. Even if input costs are rising, it is possible for treatment episode costs to fall if there are changes in the composition of treatment. A number of studies found that in recent years a considerable shift has taken place in the composition of treatment for depression (Berndt *et al.*, 2002; Berndt *et al.*, 2001), schizophrenia (Frank *et al.*, 2003), heart attacks (Cutler *et al.*, 2001) and cataract surgery (Shapiro *et al.*, 2001). Such improvements are likely to lead to an endogenous change in the patient mix as the cost of the treatment changes.

A cost-of-living approach

As an alternative to the conventional 'service price index' (SPI[4]) approach, Berndt *et al.* (2000) and Cutler *et al.* (2001) (BC) outline a cost-of-living approach, based on patient welfare[5]. Whereas the SPI approach measures changes in the price of a representative bundle of medical services over time by calculating the amount of money required to purchase a particular bundle of goods at different points in time,

[4] The SPI terminology is due to Cutler *et al* (2001)
[5] Whereas Cutler *et al* (2001) are concerned with the specific treatment of heart attacks, Berndt *et al* (2000) are concerned with the medical care sector as a whole.

Berndt *et al.* (2000) and Cutler *et al.* (2001) derive their indices from welfare based concepts. Whilst SPIs do capture some essential elements of patient welfare, they may only partially account for changes in the quality of services (i.e. if the same number of units of the service produce greater utility).

Berndt *et al.* (2000) argue that there are five reasons why measurement is more difficult in the health sector than in other service sectors. The first is the problem of moral hazard, which leads to a divergence between marginal private and social costs. The second is the existence of asymmetric information, which may lead to a misalignment of the interests of patients and physicians (the so called 'principal-agent' problem). Third, it is difficult to evaluate the widespread technological and organisational changes that result in the introduction of new services. Fourth, the market structure of the US health care sector involves licensing, regulation and intellectual property rights resulting in market power in the hands of suppliers and the involvement of government and large insurance companies leading to market power in the hands of buyers. Finally, it is the fact that medical services are intermediate goods in the production of health that makes measurement of the output (and prices) of the medical sector particularly difficult.

In the BC approach, consumers purchase goods and services to maximise some notion of utility where most goods and services give direct benefits (final purchases such as cars, food, services of hairdressers etc) but medical services are intermediate goods that increase utility indirectly. Consumers do not value these services *per se*, but rather they value the resulting health outcomes. Therefore, it is the impacts on health that are conceptually the composite good that should be priced. Medical services are only one of a number of factors that enter the production function for health. Others include knowledge of how to use medical treatments[6], time spent in seeking and receiving treatments and in recovery[7], lifestyles (e.g. eating and drinking habits, smoking, exercise) and the environment (new diseases, pollution).

[6] Knowledge can be seen as disembodied technological progress distinct from new technologies embodied in new treatment methods and associated inputs. This knowledge has public good attributes and so it is difficult to assess the impacts of changes in knowledge on prices of suppliers versus those facing consumers.

[7] For example lowering the recovery time will reduce the cost to consumers but have little or no impact

The individual's utility function will include health and many other arguments, such as their lifestyle, leisure time, consumption of goods and services, the environment etc. Many of the arguments appearing in the utility function also appear in the health production function so that these have both direct and indirect impacts. An important implication is that the output of the medical care industry should be seen as the *marginal* impact of health on utility, holding constant other inputs affecting health (such as lifestyle), and should not be envisaged as the *average* health of the population. These points are also emphasised by Triplett (1999, 2001).

Consider the case of a representative consumer choosing between consumption of goods and services (other than medical services) and health, where for simplicity they assume there is only one disease which everyone contracts[8]. Letting Y denote exogenous income, H the health state, M and P_M the quantity and price of medical care, I and P_I as the quantity and price of a constant-quality insurance policy, K as medical knowledge, E the environment, L as leisure time and T_M as the time allocated to medical treatments. The utility function is written as:

$$U = U\left[Y - P_M M - P_I I,\ H(M,K,E),\ L - T_M\right] \tag{50}$$

The first term is non-medical care consumption, the second is health and the third is non-medical care time. Note that the equation makes no assumptions about how medical treatment prices are set or decisions about medical treatments.

The question posed is: assuming the individual maximises their utility at each time period, what is the correct price index for changes between periods 0 and 1? We can define Laspeyres (Paasche) index as the change in funds the individual needs in period 1 (0) to be as well off as in period 0 (1)[9]. In the Laspeyres case, the additional amount C is defined by:

$$\begin{aligned} &U\left[Y - P_{M1}M_1 - P_{I1}I_1 + C,\ H\left(M_1,K_1,E_1\right),\ L - T_{M1}\right] \\ &= U\left[Y - P_{M0}M_0 - P_{I0}I_0,\ H\left(M_0,K_0,E_0\right),\ L - T_{M0}\right] \end{aligned} \tag{51}$$

---

for producers.

[8] Cutler *et al* (2001) extend this to the case of multiple diseases with consumers having probabilities of contracting various illnesses.

[9] Alternatively, we could define the Fisher index, which is the geometric mean of the two.

C is the change in the cost of living. To form a price index one can scale C by the income required to produce utility Y in period 0, i.e.

$$\text{(52)} \qquad \text{Cost of Living Index } = 1 + \frac{C}{Y}$$

Using a first order difference approximation, and letting X denote non-medical consumption ($Y - P_M M - P_I I$), we can differentiate and rearrange (51) to give:

$$\text{(53)} \qquad C \cong \frac{d\left(P_M M + P_I I\right)}{dt} - \frac{U_H}{U_X}\left\{H_M\left(\frac{dM}{dt}\right) + H_K\left(\frac{dK}{dt}\right) + H_E\left(\frac{dE}{dt}\right)\right\} + \frac{U_L}{U_X}\left(\frac{dT_M}{dt}\right)$$

where U_H is the marginal utility of health, U_X is the marginal utility of non-medical consumption, U_L is the marginal utility of leisure, assuming $dC/dt = U_X$. We can see from (53) that the change in the cost of living is made up of three parts: the additional spending on medical care and insurance; the dollar value of the change in health over time; and the change in the time cost of receiving medical care.

The additional spending on medical care and insurance over time may arise due to increased quantities or prices. Increases in costs of medical services raise the cost of living index. Note that if the medical environment changes, e.g. a new disease appears, medical expenses will probably increase but this should not be viewed as a change in the cost of living, since the latter's computation assumes an unchanging environment. Similarly, since outcomes are being held fixed, if one treatment becomes ineffective (e.g. antibiotics become less effective with new strains of bacteria) and is replaced by a more expensive treatment (drugs), the price index should increase reflecting the lower efficacy (quality deterioration) of the original drugs.

An improvement of health lowers the cost of living through any of the three channels (changes in the quantity of medical services, knowledge or the environment). The money value of this change is calculated using the marginal rate of substitution between health and other goods $-(U_H/U_X)$. Similarly, the change in the time of receiving medical care is converted into money by multiplying by the marginal rate of substitution between leisure and other goods (usually the after-tax wage). If the use of a different treatment reduces recovery time then the cost of living falls. Aggregation

to many consumers can be achieved using expenditure weights applied to each consumer's utility. There may also be issues relating to the differences between the cost of living a year and the cost of living a lifetime, e.g. if a new discovery vastly raises life expectancy.

A fundamental issue is how to estimate the values of the variables in equation (52) and (53). Four alternatives are: (*i*) hedonic analysis to separate the value of services to the patient from pure price effects, as in studies for mental health discussed below; (*ii*) hedonic regressions based on insurance policies, in combination with willingness to pay techniques (e.g. Pauly, 1999); (*iii*) to make specific assumptions on the way in which medical treatment decisions are made (e.g. Cockburn and Anis (2001) for prescription drugs); (*iv*) a more direct measurement method that focuses on a particular disease and estimates empirically the changes in treatment costs and medical outcomes for that disease. This final example is the method employed in Cutler *et al.* (2001) for heart attacks. Berndt *et al.* (2000) also argue for a medical care expenditure price index based on episode treatment costs, a suggestion made by Anne Scitovsky back in 1962 (Scitovsky, 1964; Scitovsky, 1967).

Cutler *et al.* (2001) contrast input price indices for the cost of heart attack treatment with an outcomes-adjusted index that takes account of changing treatment regimes and incorporates a valuation of the extension of life expectancy attributable to new heart attack treatments. They attempt to measure a Laspeyres version of (52) (i.e. using income in the initial period as the denominator), measuring quality changes via mortality rates[10]. By making an assumption on the dollar worth of health improvements from the literature, they are able to calculate the various individual factors in this cost of living index.

Table D.1 contrasts the service price index for heart attacks with the cost of living index, annual percentage changes from 1984 to 1994 adjusted for general inflation. The results are shown using Medicare data. The disaggregated SPI is based on sampling services produced. The estimates in the Table illustrate the fixed weight bias problem – this is mainly due to shorter average lengths of stay in hospital together

[10] They also present a variant that uses QALYs rather than mortality rates as the outcome measure.

with large increases in the price of a hospital day. The aggregated price index uses the same service information but aggregates all components for each of four different treatment regimes. This lowers the fixed weight index by 0.5% per annum, but the difference between the fixed weight and annual chain linked index is much less than for the disaggregated case. Most striking in the table is the difference between the SPI and the base case cost-of-living index. The authors interpret these results as pointing to a large bias from a failure to take account of outcome changes. Although the range in the COL indices is very large[11], the conclusion remains that there is a substantial upward bias in the conventional price indices.

Table D.1 Price changes for heart attack treatments

	Real Annual % change
*Disaggregated price index**	
Fixed basket index	2.8
Five-year chain index	2.1
Annual chain index	0.7
Aggregated service price index	
Fixed basket index	2.3
Annual chain index	1.7
Cost of living index	
Years of Life [range**]	-1.5 [-0.2, -13.7]
Quality of Life [range**]	-1.7 [-0.3, -16.8]

Source: Cutler et al. Table 8.4, p.321
* includes data for a major teaching hospital as well as Medicare;
** based on higher and lower estimates of the net value of a life year.

Mental health

Mental health is a particularly interesting area for investigation as there has been particular concern over the increase in expenditure for this type of care, with many official statistics reporting price increases for mental health treatment. It is also a particularly difficult type of care to evaluate in terms of outcomes and outputs. What is clear from the work is that it is important to focus on the direct medical costs of

[11] The direct approach to measuring a cost of living index is sensitive to the values placed on an additional year of life and represents one extreme position in terms of adjusting for quality changes.

treating an episode of an illness, rather than changes in the prices of the inputs used in treatment. This is because even if input costs are rising, it is possible for treatment episode costs to fall if there are changes in the composition of treatment (Berndt, 2003). For example, a number of studies conducted by Berndt, Busch and Frank found a considerable shift in the composition of treatment for depression has taken place in recent years, with the combination of psychotherapy and TCAs being replaced by the use of SSRI drugs, sometimes also in combination with therapy. It is also clear that there needs to be some adjustment for quality.

Rising expenditure on mental health has generated considerable interest in constructing price indices for this area of medical care, in particular: major depression, schizophrenia and bipolar disorder. For all three disorders, studies suggest that the price of treating an episode or individual have declined in recent years. This is contrary to many of the officially reported figures. Berndt (2003) argues that this can be explained by the fact that official statistics do not make allowances for changes in the composition of treatment over time. This suggests that the gains from both pharmaceutical innovations and institutional changes have led to treatment for a greater number of episodes and individuals (Berndt, 2003). While this is an encouraging finding, uncertainty remains as to how quality of care has been affected. The importance of quality of care is also important, but adequate data on quality and patient outcomes is not available from the medical claims data employed in most studies. Therefore, alternative approaches to accounting for quality have to be applied.

Depression

Depression accounts for approximately half of mental health expenditure in the US (Berndt, Busch and Frank, 2000/01). There have been a number of studies investigating the pricing of its treatment. Although they do not use identical methods, there is a considerable degree of similarity in both the methods employed and the conclusions reached. The studies have focused on the episode of illness, which involves pooling a number of treatments into what Frank, Busch and Berndt (1998) term ‘bundles’.

Frank *et al.* (1998) use the results from a range of clinical trials and medical literature to form bundles of treatments that are *ex ante* anticipated to lead to similar outcomes.

These are used to construct demand and supply price indices for the treatment of major depression based on the cost of an episode of illness, for the period 1991-1995. Frank *et al.* identified nine major types of treatment in retrospective medical claims data from four large self-insured employers. Closely related treatment bundles were aggregated, resulting in the use of five bundles. It is important to consider the substitutability of the treatment bundles. Fixed weight indices such as the base-period Laspeyre index and the end-period Paasche index, assume that there is no possibility of substitution among treatments. Therefore, in addition to these, the authors considered a Cobb-Douglas index (which assumes that the elasticity of substitution is unity) and a Törnqvist index (which employs average shares across adjacent time periods and hence makes no *a priori* assumption about substitutability).

All four indices showed that both the demand and supply price of treating depression had fallen between 28 and 32% between 1991 and 1995. The reported decline for the cost of treating depression is contrary to the increase in the BLS price indices over the same period. The average annual differential in price changes of approximately 15% - three times the differential found by Cutler, McClellan and Newhouse (1999) for heart attack treatment. Frank *et al.* conclude that employing standard price indices – which do not allow for changes observed in treatment composition – may lead to changes in quantity being misinterpreted as changes in price, which is a cause for concern.

Several extensions to this work have since been carried out by Berndt, Busch and Frank and other authors (for example, Berndt *et al.*, 2002; Berndt *et al.*, 2001). Berndt *et al.* (2001) included a greater number of episodes and increased the number of treatment bundles to seven. Using a Laspeyres price index, this resulted in a decline of only 0.6% per annum, compared to the 9.1% per annum decline reported by Frank *et al.* (1998).

Berndt *et al.* (2002) add to the previous research by using a method that combines clinical evidence and expert opinion. This approach has two advantages: First, it enables the calculation of treatment price indices that include variations over time in the proportion of 'off-frontier' production, which increases the number of observed episodes. Second, it means that allowance can be made for different expected outcomes, and even for those patients put on the 'waiting list' or no treatment option,

who often experience remission regardless. By doing so, Berndt *et al.* (2002) are able to assess the incremental outcome gains as a result of treatment compared to no treatment outcomes.

Using similar retrospective medical claims data (for 1991-6), the authors identified and classified episodes of acute phase depression according to treatment and patient type. Information on expected treatment outcomes was gathered using a two-stage modified Delphi procedure. Ten panellists were asked to provide opinions on expected treatment outcomes, using the Hamilton Depression Rating Scale (HDRS) score (a standard clinical measure for this illness). For each of the 120 patient-treatment cells, a literature review of clinical research was provided. The experts were asked to consider a group of 100 patients and provide an estimate of how many patients would fall into four given categories: 'no depression', 'mildly depressed', 'moderately depressed' or 'no change'. However, for 39% of the cells, substantial disagreement among the experts was noted. Each expert was then shown the group's rating for those cases where there was disagreement, and after meetings they were given the option to revise their own rating. After this process, no substantial disagreement remained.

The framework under which expected outcomes data were obtained allowed Berndt *et al.* (2002) to consider results for both full and partial remission. The probability of a full remission varies from 0.15 to 0.35, the probability of partial remission is higher. One particularly interesting finding is that approximately 40% of the treatment episodes make use of medical treatment with no or little effectiveness over the no treatment option, accounting for 20% of total spending on treatments.

Average spending per full or partial remission varies according to the treatment of those cases where there is no expected outcome rating. If these cases are excluded, or assigned the median outcome, average spending remains unchanged in 1996 from 1991. If they are assigned the worst possible outcome, average spending falls to 84 for full remission. For partial remission, using the median outcome also results in a decline in average spending to 90 in 1996. However, expenditure per incremental expected full or partial remission is preferred. Generally, the expenditure indices per incremental remission fall more rapidly than average expenditure.

A base-period Laspeyre, an end-period Paasche and a Fisher Ideal price index were constructed. The treatment price index falls for all three, but if incremental remission is considered all three show a small rise. The trends in price indices are slightly sensitive to how expected outcomes are treated. It may be that there have been changes in the patient population and treatment bundles.

To account for this potential change in patient mix, hedonic equations were estimated. Eight patient categories were identified, depending on the presence of medical comorbidity, gender, age if female and comorbid substance abuse. Three alternative hedonic equations for the price per expected full remission were estimated by OLS. The results suggest that patient categories have significant and sizeable effects on treatment costs; comorbid substance abuse and comorbid medical conditions are particularly likely to increase treatment costs. Also, treatments having higher probability of full remission have higher costs *ceteris paribus*. All three equations estimate reductions in price. The results for partial remissions lead to similar findings. Estimating hedonic equations produces different results to those shown by the price indices, a reflection of the changing mix of patients over time. Overall, the cost of treating episodes of depression, conditional on expected outcome, has generally declined between 1991 and 1996. The results of the hedonic regressions indicate a fall of between -1.66% and -2.13% per annum over this period.

Berndt *et al.* (2002) conclude that using expected outcomes data gained in this way can help to construct price indices for medical treatments that allow for variation in expected outcomes and patient populations. However, it is acknowledged that the reliability of the outcome measures are a key determinant of the usefulness of this method, and validity checks are carried out to assess this. In conclusion, since expenditures on depression are thought to be increasing over this period, it seems that the source of this is an increase in volume, as the cost of treating an episode of depression seems to have fallen.

Schizophrenia

This approach of pricing an episode of care has also been implemented for other

mental health disorders. Frank *et al.* (2003) investigate the changing costs of treating schizophrenia. Using data from two counties in Florida over the period 1994/5 to 1999/00, this study sets out to answer three main questions: what is the cost of treating an individual with schizophrenia on a per annum basis?; how have composition and quality of care altered over time?; and how are estimates of changes in treatment costs affected if quality is controlled for? As the 1990s saw the introduction of significant new treatments, Frank *et al.* suggest that considerable changes in treatment patterns were likely to have taken place, impacting in turn on both the price and quality of care.

Producer price indices are constructed based on annual episodes of care. As the nature of schizophrenia is such that it would be considerably more complex for clinical experts to estimate probable outcomes, Frank *et al.* use the recommendations made by the Patient Outcomes Research Team (PORT) to account for quality. Five of their recommendations referred to types of treatment observable from the claims data – the use of pharmacotherapy was found to be beneficial for example, whereas the evidence for vocational rehabilitation appeared to have little effect. Frank *et al.* then used these to construct treatment bundles, four of which consisted of single treatments, such as any antipsychotic medication and others which comprised more than one form of treatment such as medication and therapy. An 'all other treatment' category was also formed. Each person-year observation was then categorised to one or more of the treatment bundles as appropriate.

The authors note that a shortcoming of measuring quality in this way is that it merely reveals whether a patient has received treatment, and provides no indication of whether the amount of treatment was sufficient. They also point out that in some sense this reflects only minimum standards of care, as it seems reasonable to expect that a patient should receive at least one form of treatment.

Multivariate regression equations were estimated, using the natural log of annual mental health direct medical costs as the dependent variable. Explanatory variables included patient characteristics, enrolment and medical histories and indicator variables for the treatment bundles. As this period saw significant changes in treatment, Chow tests were applied to test for parameter stability. This resulted in the

rejection of parameter equality, and therefore the data were not pooled across years. Quality and patient characteristics were held constant at a point in time, and predicted spending compared. This raises further questions over which year should be used to fix these factors, and various options are explored. An important finding is that, contrary to what one would normally expect, the fixed weight and chained Laspeyre price indices fall more rapidly than the comparative Paasche indices. Frank *et al.* interpret this as reflecting a rightward shift of the demand curve, as over time doctors accumulate knowledge of the efficacy of the newer pharmaceuticals, and alter their prescribing behaviour leading to positive correlations between price and quantity.

Frank *et al.* reveal that there have been significant changes in treatments given to patients with schizophrenia. A considerable increase in atypical antipsychotics has been observed, replacing older pharmaceutical treatments, in line with PORT guidance. At the same time, there has been a decline in the use of therapy and psychosocial rehabilitation. This leads Frank *et al.* to raise the concern that a significant amount of treatment is not supported by clinical evidence.

The Frank *et al.* study demonstrates the considerable importance of controlling for quality. Without any adjustment for quality, mean mental health-related treatment costs for individuals with schizophrenia increased 2.4% over the period. When treatment quality is held constant, both the fixed weight and chained Fisher Ideal indices showed cumulative price falls of more than 22% over the period. The results show that without adjusting for quality changes, there has been a rise of 0.5% per annum in the individual annualised cost of treating schizophrenia. If quality of care and patient characteristics are held constant over time, mean treatment costs have fallen by 5.5% per annum between 1994/5 and 1999/2000, a cumulative price decline of around 25% over the period, with the greatest decline occurring in the first half of the 6-year period under consideration. Overall, they conclude that as the cost of treating an individual per annum has declined, the observed increase in overall expenditure is an indication that there has been an increase in the number of individuals being treated.

Bipolar disorder

Research by Ling *et al.* (2002), discussed in Berndt (2003) investigates the changing costs of treating bipolar I disorder. Like schizophrenia, there is at present no known cure for bipolar I disorder, so the focus of medical care in this case is to try and improve symptoms, including the prevention of recurrences. In contrast however, there has not been such a dramatic change in the composition of treatment.

Ling *et al.* (2002) estimate a multivariate regression model, similar to that used by Frank *et al.* (2003), based on five mutually exclusive and exhaustive treatment bundles. The results suggest a fall of 31% in the cost index between 1991 and 1995, compared to a decline of 12% if no allowance is made for changes in patient mix and treatment composition.

Cataract surgery

We noted earlier that cataract surgery has undergone considerable advances in recent decades. As a result, the number of cataract operations performed has risen significantly and cataract patients increasingly receive surgery at an earlier stage of the disease. Shapiro *et al.* (2001) identify the good to be valued in the case of cataract surgery as the visual improvement resulting from the intervention. The benefit of the treatment is equal to the difference between the visual function following surgery and the visual function without surgery. This varies for serious, progressive or mild cases of disease. This framework allows incorporation of the lifetime gains from treatment and account to be taken for endogenous change in the patient mix as the cost of the treatment changes. Nevertheless the benefit still needs to be valued. QALYs may be used, although again this requires a value to be placed upon the QALY. Willingness to pay (WTP) methods could be used to value either a QALY or the cataract surgery directly. However, even after this value is determined, other problems still remain, such as the unknown population distribution of the course of the disease.

Shapiro *et al.* (2001) construct two price indices for measuring the monetary cost of cataract surgery: a hypothetical CPI is calculated and a prototypical cost index. As for

a typical CPI, weights are applied to the price indices for the individual inputs reflecting their relative importance; in this case, the inputs are physician and hospital services. Hypothetical benchmark years of 1969 and 1979 are set, and the component indices are then weighted by the shares in these years. Shapiro *et al.* note that a major disadvantage of this method is that technical progress is only reflected in the rate of growth of the index, reflecting the change in the component price indices, not the index level. In cataract surgery therefore, where there has been a significant fall in the hospital services required for treating a patient, this will not be reflected in the level of the hypothetical CPI.

The prototypical cost index was calculated by $\sum_i \left(q_{it} p_{it} / q_{ib} p_{ib} \right)$, where q refers to the quantity of each input i, p refers to the CPI for input i, and b indicates the base year. An advantage of this index over the hypothetical CPI is that it is able to incorporate changes in the level of inputs, and therefore reflect the decline in the quantity of hospital services used.

The results show that the prototypical index grows much less rapidly than the hypothetical CPI. Over the period 1969 to 1994, the prototypical index increased by 4.1% per annum, compared to growth of 9.6% per annum in the hypothetical CPI. Compared to the total CPI, taken over the whole period, the prototypical index actually declines. Therefore, the monetary cost of treating cataracts has actually been increasing at a slower rate than the general price level over this period even without making an adjustment for improvements in quality or the value of undergoing surgery at an earlier stage of the disease. Shapiro *et al.* highlight that these are also important factors that need to be accounted for when constructing price indices.

Shapiro *et al.* make three main recommendations based on their study: First, they advocate that the cost of medical care should be measured in terms of the unit values for treating a condition, classifying the good as the treatment of the illness and the measuring cost in terms of the current inputs required for treatment[12]; Second, they note that the durability of the benefits resulting from treatment should be accounted for. The problem of durability is an issue for many goods, not just those in medical care. Cataract surgery leads to benefits that last over many periods, and accounting for

[12] Unfortunately, data limitations prevent the adoption of a unit value approach for cataract surgery

this characteristic is critical for correctly accounting for the procedure in a price index. The future value of not having to endure a period of deteriorating quality of life needs to be incorporated.

Lastly, an allowance needs to be made for the endogenous change in patient mix as a result of changing cost. The value of a patient receiving treatment sooner or receiving treatment when previously they would not have done at all is extremely important. This is further complicated by the fact that different individuals receiving surgery place very diverse valuations upon it, some of which are impossible to know at the time of intervention. For example, how the disease would have progressed without treatment will affect this valuation. The authors note that this could have significant consequences for valuing the benefit of the treatment.

Shapiro, Shapiro and Wilcox (1999) discuss how prices and output in medical care generally may be more accurately measured. A framework is developed for evaluating demand for medical care, allowing for certain characteristics of the health care sector that make such measurement particularly complex, namely the durability of medical treatment, technical progress and third-party payment.

First, as highlighted in Shapiro *et al.* (2001), the benefits of a medical intervention affect the utility of the patient not only in the period immediately after treatment, but for the remainder of their life. This durability needs to be considered. Secondly, the use of third-party payers in health care means that the monetary cost of care often has little impact on the patient's demand for treatment. However, costs also exist in terms of the burden to the patient, and these will impact upon their demand for care. Third, technical change can lead to a considerable improvement in outcomes, and also reduce cost in terms of patient burden. The authors note that given the pace of technical progress in health care, this can actually create a desire to postpone treatment to benefit from future advances. Monetary costs may increase or decrease as a result of technical progress. Shapiro *et al.* note that when technology results in an increase in the treatment value, more treatment is likely to take place, whether this involves the treatment of a greater number of patients and/or intervention at an earlier

---

similar to that in Cutler *et al.* (2001).

stage of the illness. The authors report that such a response has been identified for several conditions, including cataracts, angioplasty and joint replacement. Interestingly, it is highlighted that an increasing number of interventions may not always be beneficial – for example, Chernew, Fendrick and Hirth (1997) suggest that, in the case of gall bladders, falling patient burden has resulted in some unwarranted surgery.

Using this framework for analysing demand, Shapiro *et al.* propose how health care price indices may be adjusted for quality improvements. Improvements in outcomes and reduced patient burden should be reflected in the price index. The authors argue that to account for such quality changes, the net benefit of the treatment needs to be measured, and changes in this net benefit should then be used to adjust the unit cost. Several methods for measuring net benefit are discussed. Willingness to pay is one possibility, but incurs problems with data availability. Alternatively, patient surveys may help to value these benefits. Such surveys may use contingent valuation methods, where patients are asked about hypothetical health outcomes, or QALYs. An important contribution to the valuing process can also be made by experts in the field. Despite criticisms of survey methods, in reality these are the most probable approach given the lack of data for willingness to pay methods. Finally, Shapiro *et al.* emphasise that allowance must also be made for the uncertainty and variation involved in terms of both the success of the intervention and how the illness would have progressed without treatment.

This leads Shapiro *et al.* to make a number of recommendations for price measurement in health care. They argue that a 'standard metric' for quality-adjusting the price of treatments is required, incorporating the knowledge of medical professionals and experts. They also propose a two-step approach to developing a 'database of values of health care outcomes'. As improvements in outcomes occur, this could be used to adjust the price of interventions. This would first involve the determination of values for 'broad dimensions of impairments', such as blindness, deafness and mortality, followed by an ongoing evaluation of how changes in treatments impact upon outcomes in terms of these dimensions. In this manner, consistency across varying diagnoses could be attained.

Conclusions

What emerges from the US price literature is that it is important to focus on the direct medical costs of treating an episode of an illness, rather than changes in the prices of the inputs used in treatment, and that there needs to be some adjustment for quality. Even if input costs are rising, it is possible for treatment episode costs to fall if there are changes in the composition of treatment. Such improvements are likely to lead to an endogenous change in the patient mix as the cost of the treatment changes.

The conclusion from the mental health and other studies is that it is not necessary to go to the extreme of valuing lives saved in order to adjust for quality change. However the cost of this approach is that the methods may only partially capture improvements in outcomes due to changes in medical treatments. Triplett (1999) argues that the outcomes studies based on cost-effectiveness research can be used to identify medical care outputs that adjust for quality change. His suggestion is to combine information on prices and quantities and outcome measures such as QALYs to derive price indices that take account of new treatments with different QALY implications. Changes in QALYs for existing treatments can readily be incorporated into this framework.

E Data sources and data availability

There are a variety of data sources that we are currently exploring both for activity data as well as data on inputs. While these will be expanded upon in Phase 2 of this project, some initial descriptions of some of the main data sources are given here.

There is reasonably comprehensive coverage by routine administrative data on hospital activities (via Hospital Episode Statistics), and also on community health service activities via the Korner returns (though the quality of the latter is variable). Together these account for 64% of NHS expenditure. There is much poorer coverage of family health services, especially of primary care services, which account for 24.5% of NHS expenditure. The remaining 12.5% covers activity such as national screening programmes, R&D, etc., where data are patchy. A tendency to shift activity away from hospitals into general practice and community settings where data are less readily available means that estimates of NHS productivity growth may be biased downwards (Yuen and Towse, 2003).

Hospital Episodes Statistics (HES) data description and data availability

Variable	Description	Key	1996/97	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03
hesid	uniquely identifies patient across all data years, generated using a combination of NHS number, local patient identifier, patients' postcode, sex and date of birth		y	y	y	y	y	y	y
epikey	record identifier created by the system		y	y	y	y	y	y	y
admidate	date the patient was admitted to hospital at start of spell, recorded on first and all subsequent episodes within the spell		y	y	y	y	y	y	y
admimeth	code which identifies how patient was admitted to hospital, recorded on first and all subsequent episodes within the spell	11 Elective - from waiting list 12 Elective - booked	y	y	y	y	y	y	y

		13	Elective - planned							
		21	Emergency - via A&E, including casualty							
		22	Emergency - via GP							
		23	Emergency - via Bed Bureau, including Central Bureau							
		24	Emergency - via consultant out-patient clinic							
		28	Emergency - other means, including A&E of another provider							
		31	Maternity - baby delivered after mother's admission							
		32	Maternity - baby delivered before mother's admission							
		81	Transfer from another provider other than an emergency							
		82	Other - babies born in provider							
		83	Other - babies born outside provider, except when born at home as intended							
		84	Admission by admission panel of High Security Psychiatric Hospital (HSPH)							
		89	From admissions waiting list of HSPH							
		98	Not applicable (e.g. other maternity event)							
		99	Not known							
admisorc	code which identifies where patient was immediately prior to admission	19	The usual place of residence, including no fixed abode	y	y	y	y	y	y	y
		29	Temporary place of residence when usually resident elsewhere (e.g. hotels, educational establishments)							
		30	Repatriation from HSPH (from 1999-00)							
		37	Penal establishment - court (from 1999-00)							
		38	Penal establishment - police station (from 1999-00)							
		39	Penal establishment (court and police station excluded from 1999-00)							
		48	HSPH, Scotland (from 1999-00)							
		49	Special HA establishment under the HSP Services Commissioning Board							
		50	NHS provider - medium secure unit (from 1999-00)							
		51	NHS provider where the patient was in general ward or A&E							
		52	NHS provider where patient was in maternity or neonatal ward							
		53	NHS provider where the patient was in a ward for people who are mentally ill or who have learning disabilities							

		54 NHS run nursing home, residential care or group home 65 Local authority Part 3 residential accommodation - where care is provided (from 1996-97) 66 Local authority foster care, but not in Part 3 residential accommodation - where care is provided (from 1996-97) 69 LA home or care (1989-90 to 1995-96) 79 Babies born in or on the way to hospital 85 Non-NHS (other than Local Authority) run residential care home (from 1996-97) 86 Non-NHS (other than Local Authority) run nursing home (from 1996-97) 87 Non-NHS run hospital 88 Non-NHS (other than Local Authority) run hospice 89 Non-NHS institution (1989-90 to 1995-96) 98 Not applicable 99 Not known							
category	administrative and legal status of a patient	10 NHS patient - not formally detained 11 NHS patient - formally detained under Part II of Mental Health Act 1983 12 NHS patient - formally detained under Part III of Mental Health Act 1983 or under other Acts 13 NHS patient - formally detained under part X, Mental Health Act 1983 * 20 Private patient - not formally detained 21 Private patient - formally detained under Part II of Mental Health Act 1983 22 Private patient - formally detained under Part III of Mental Health Act 1983 or under other Acts 23 Private patient - formally detained under part X, Mental health Act 1983* 30 Amenity patient - not formally detained 31 Amenity patient - formally detained under Part II of Mental Health Act 1983 32 Amenity patient - formally detained under Part III of Mental Health Act 1983 or under other Acts	y	y	y	y	y	y	n

		33	Amenity patient - formally detained under part X, Mental health Act 1983 *							
classpat	patient classification	1	Ordinary admission	y	y	y	y	y	y	y
		2	Day case admission							
		3	Regular day attender (from 2002-03)							
		4	Regular night attender (from 2002-03)							
		5	Mothers and babies using only delivery facilities							
		8	Not applicable (other maternity event) from 1996-97							
		9	Not known 1996-97 to 1999-00							
admincat	administrative category on admission	01	NHS patient, including overseas visitors charged under Section 121 of the NHS Act 1977 as amended by Section 7(12) and (14) of the Health and Medicine Act 1988	y	y	y	y	y	y	y
		02	Private patient, who uses accommodation or services authorised under section 65 and/or 66 of the NHS Act 1977 (Section 7(10) of Health and Medicine Act 1988 refers) as amended by Section 26 of the National Health Service and Community Care Act 1990							
		03	Amenity patient, who pays for use of single room or small ward in accord with section 12 of the NHS Act 1977, as amended by section 7(12) and (14) of the Health and Medicine Act 1988							
		04	A category II patient, for whom work is undertaken by medical or dental staff in categories II paragraph 37 of Terms and Conditions of Service of Hospital Medical and Dental Staff							
		98	Not applicable							
		99	Not known							
diag_1	diagnosis code, diag_1 to diag_3 contain information about a patient's illness or condition, diag_1 contains primary diagnosis, other fields contain secondary diagnoses, defined in International Statistical Classification of Diseases, Injuries and Causes of Death (ICD-10)			y	y	y	y	y	y	y
diag_2				y	y	y	y	y	y	y
diag_3				y	y	y	y	y	y	y
oper_1	operation, field oper_1 contains main (most resource intensive) procedure, other field secondary procedures, defined in Tabular List of the Classification of Surgical Operations and Procedures, version OPCS4			y	y	y	y	y	y	y
oper_2				y	y	y	y	y	y	y

hrgorig	Healthcare Resource Group (HRG) initially calculated for hospital episode		y	y	y	y	y	y	n
hrglate	Healthcare Resource Group (HRG) most recently calculated for a hospital episode		y	y	y	y	y	y	y
mainspef	main specialty under which consultant is contracted	100 General surgery 101 Urology 110 Trauma and orthopaedics 120 Ear, nose and throat (ENT) 130 Ophthalmology 140 Oral surgery 141 Restorative dentistry 142 Paediatric dentistry (available from 1999-00) 143 Orthodontics 150 Neurosurgery 160 Plastic surgery 170 Cardiothoracic surgery 171 Paediatric surgery 180 Accident and emergency (A&E) 190 Anaesthetics 191 Pain management (available from 1998-99) 300 General medicine 301 Gastroenterology 302 Endocrinology 303 Haematology (clinical) 304 Clinical physiology 305 Clinical pharmacology 310 Audiological medicine 311 Clinical genetics 312 Clinical cytogenics and molecular genetics (available from 1990-91) 313 Clinical immunology and allergy (available from 1991-92) 314 Rehabilitation (available from 1991-92) 315 Palliative medicine 320 Cardiology 330 Dermatology	y	y	y	y	y	y	y

340	Thoracic medicine
350	Infectious diseases
360	Genito-urinary medicine
361	Nephrology
370	Medical oncology
371	Nuclear medicine
400	Neurology
401	Clinical neuro-physiology
410	Rheumatology
420	Paediatrics
421	Paediatric neurology
430	Geriatric medicine
450	Dental medicine (available from 1990-91)
460	Medical opthalmology (available from 1993-94)
501	Obstetrics for patients using a hospital bed or delivery facilities
502	Gynaecology
560	Midwifery (available from October 1995)
610	General practice with maternity function
620	General practice other than maternity
700	Learning disability (previously known as mental handicap)
710	Mental illness
711	Child and adolescent psychiatry
712	Forensic psychiatry
713	Psychotherapy
715	Old age psychiatry (available from 1990-91)
800	Clinical oncology (previously Radiotherapy)
810	Radiology
820	General pathology
821	Blood transfusion
822	Chemical pathology
823	Haematology
824	Histopathology
830	Immunopathology
831	Medical microbiology

		832	Neuropathology							
		900	Community medicine							
		901	Occupational medicine							
		950	Nursing episode (available from 2002-03)							
		Spaces	Other maternity event							
		&	Not known							
elecdur	waiting time in days, difference between date on which it was decided to admit patient and admission date, only applicable where elective admission (admimeth - is 11, 12, 13) was scheduled and took place			y	y	y	y	y	y	y
epidur	episode duration in days, difference between episode start date and episode end date			y	y	y	y	y	y	y
epiorder	the number of the episode within the current spell, all start with epiorder 01, if patient moves to care of another consultant, episode increments by 1 until patient is discharged (transfer to another Trust or PCT - first episode in new Trust will have epiorder 01)			y	y	y	y	y	y	y
disdate	date at end of spell, only present for last episode			y	y	y	y	y	y	y
disdest	discharge destination, where patient was due to go on leaving hospital		same code as admisorc, except:	y	y	y	y	y	y	y
		79	Not applicable - patient died or still birth							
dismeth	discharge method, completed for last episode in spell	1	Discharged on clinical advice or with clinical consent	y	y	y	y	y	y	y
		2	Self discharged or discharged by a relative or advocate							
		3	Discharged by a mental health review tribunal, the Home Secretary or a court							
		4	Died							
		5	Baby was still born							
		8	Not applicable - patient still in hospital							
		9	Not known: a validation error (from 1996-97)							
sex	sex of patient	1	male	y	y	y	y	y	y	y
		2	female							
		0 or 9	not known / not specified							
startage	age in whole years from episode start date and date of birth (dob), for patients under 1 year old special codes in the range 7001 to 7007 apply			y	y	y	y	y	y	y
ethnos	code which specifies ethnic groups and nationalities, from 2001 codes changed to conform to 2001 census classification		to 2000-01:	y	y	y	y	y	y	y
		0n	White							

		1n Black - Caribbean 2n Black - African 3n Black - Other 4n Indian 5n Pakistani 6n Bangladeshi 7n Chinese 8n Any other ethnic group 9n Not given Xn Not known from 2001-2002: An British (White) Bn Irish (White) Cn Any other White background Dn White and Black Caribbean (Mixed) En White and Black African (Mixed) Fn White and Asian (Mixed) Gn Any other Mixed background Hn Indian (Asian or Asian British) Jn Pakistani (Asian or Asian British) Kn Bangladeshi (Asian or Asian British) Ln Any other Asian background Mn Caribbean (Black or Black British) Nn African (Black or Black British) Pn Any other Black background Rn Chinese (other ethnic group) Sn Any other ethnic group Zn Not stated							
marstat	defines a patient's marital status	1 Single 2 Married, including separated 3 Divorced 4 Widowed 8 Not applicable 9 Not known	y	y	y	y	y	y	y

ward91	patient's full frozen 1991 Census electoral ward, derived from patient's postcode		y	y	y	y	y	y	y
res_currward	current electoral ward within a given district where the patient resided, derived from patient's postcode		y	y	y	y	y	n	y
resladst	local authority district, code for the county (first two characters) local authority district (last two characters) of residence derived from patient's postcode, used in conjunction with currward to produce unique value indicating ward within a given district where patient resided	nnnn Local authority code S Scotland U England (NOS) W Wales Y Not known Z Northern Ireland X Foreign (from 1991 onwards)	n	n	n	n	n	y	n
currward	current electoral ward of patient, derived from patient's postcode	aa Electoral ward	n	n	n	n	n	y	n
pcgcode	derived field providing the Primary Care Group responsible for the patient		n	n	n	y	y	y	n
pcgorig	basis on which the PCG code was assigned	1 GPPRAC was used to derive the code 2 REGGMP was used to derive the code 3 PURCODE was used to derive the code 4 POSTCODE was used to derive the code 5 POSTCODE allocated code, PCG code was blank 6 POSTCODE allocated code, PCG code was 49998 9 PCG code not known	n	n	n	y	y	y	n
pctcode	derived field providing the Primary Care Trust responsible for the patient		n	n	n	n	n	n	y
pctorig	basis on which the PCT code was assigned	1 GPPRAC was used to derive the code 2 REGGMP was used to derive the code 3 PURCODE was used to derive the code 4 POSTCODE was used to derive the code 5 POSTCODE allocated code, PCT code was blank 6 POSTCODE allocated code, PCT code was 59998 9 PCT code not known	n	n	n	n	n	n	y

gpprac	code which defines patient's registered GP Practice (registered GMP)	Vnnnnn GP's practice code (English GP's with codes commencing A-P only) V81998 No referring doctor, therefore no practice code V81998 MoD doctor V81999 Unknown	n	n	y	y	y	y	(not yet)
procode	provider code, procode5 contains complete NHS provider code (organisation code plus site code), procode3 contains first three characters (organisation code)		y	y	y	y	y	y	y
siteret	code that defines the site on which the patient was treated		n	n	n	n	n	n	y
purcode	purchaser code for organisation commissioning patient's treatment, six main types dependent on data year:	Health Authorities GP fund holders Primary care Groups Primary care Trusts Private patients Department of Health - for patients from abroad and certain UK residents, notably those treated by Special HAs	y	y	y	y	y	y	y
resha	code for the Health Authority (HA) in which the patient lived immediately before admission, available up to 2002-03 derived from patient's postcode, from 2002-03 information relates to Strategic Health Authority of residence		y	y	y	y	y	y	y
gross_a	coverage grossing factor, compensates for missing or additional HES records, correction factor, default is 1.0000		y	y	y	y	y	y	y
gross_b	combined grossing factor, combines administrative (coverage) grossing and clinical grossing which compensates for under counting of diagnoses or operations where cases contain not known diagnosis code, shortfall in diagnostic coding used to calculate compensating factor, multiplied by gross_a to produce gross_b, default is 1.0000		y	y	y	y	y	y	y

Reference Cost data description and data availability

Organisation	Acronym	Activity Category	Type of Activity Code	Units of Activity	2003	2002	2001	2000	1999	1998
NHS TRUSTS	**TELIP**	Elective inpatients	HRG	FCE	*	*	*	*	*	*
	TNELIP	Non-Elective inpatients	HRG	FCE	*	*	*	*	*	*
	TDC	Day cases HRG	HRG	FCE	*	*	*	*	*	-
	TCCS	Critical Care Services	Service Code	Bed Days	*	*	*	*	*	-
	TOP HRG	Outpatient HRG	HRG	Attendances	*	*	*	*	*	-
	TOPS F	Outpatient First Attendance	Speciality Code	First Attendances	*	*	*	*	-	-
	TOPS FU	Outpatient Follow up Attendance	Speciality Code	Follow up Attendances	*	*	*	*	-	-
	TOPS MAT	Outpatient Maternity Services Attendance	Service Code	First Attendances	*	*	*	-	-	-
	TOPWA	Outpatient Ward Attendance	Service Code	Ward Attenders	*	-	-	-	-	-
	TCDCWA	Ward Attenders	Service Code	Ward Attenders	*	-	-	-	-	-
	TRDNA	Regular Day and Night Admissions	Service Code	Admissions	*	-	-	-	-	-
	TAH	Audiological Services	Service Code	Hearing Aids	*	*	*	-	-	-
	TPATH	Direct Access: Pathology Services Test	Speciality Code	Tests	*	*	*	-	-	-
	TRADIO	Direct Access: Radiology Services Test	Banding Code	Tests	*	*	*	-	-	-
	TCMTEST	Direct Access: Clinical Measurement	Code	Tests	*	-	-	-	-	-
	TTPHYS (2)	Direct Access Therapy Services	Service Code	Clients Seen	*	* [(a)]	-	-	-	-
	TA&E	Accident and Emergency Services	HRG	Attendances	*	*	*	-	-	-
	TA&E (2)	Accident and Emergency Services: Minor Injury Service	HRG	Attendances	*	-	-	-	-	-
	TA&E (3)	Accident and Emergency Services: Observation/Pre-Admission/Medical Assessment Units	HRG	HRG	*	-	-	-	-	-
	TRDTHY	Radiotherapy Treatments/Fractions	HRG	Treatments	*	*	*	*	-	-

TCHEMO	Chemotherapy	Service Code	Regimens	*	*	-	-	-	-
TRENALD	Renal Dialysis Sessions	Service Code	Sessions	*	*	*	-	-	-
TTPT	Renal Transplant Episode	Service Code	Episodes	*	-	-	-	-	-
TCF	Cystic Fibrosis Inpatient	Service Code	Various	*	-	-	-	-	-
TBMT	Bone Marrow Transplant	Service Code	Episodes	*	*	*	-	-	-
TSI	Spinal Injuries Unit Occupied Bed Days	Service Code	Occupied Bed Days	*	*	-	-	-	-
TREHAB	Rehabilitation Services Occupied Bed Days	Service Code	Occupied Bed Days	*	*	-	-	-	-
TDCF	Day Care Facilities Regular Attendance	Service Code	Patient Days	*	-	-	-	-	-
THAH	Hospital at Home Team Client	Service Code	Team Clients Seen	*	-	-	-	-	-
TCNS	Community Nursing Services	Service Code	Various	*	-	-	-	-	-
TCNS (2)	Community/Outreach Specialist Nursing Services	Service Code	Clients Seen	*	* [b]	* [b]	-	-	-
TCMS	Community Midwifery Services	HRG	Numbers	*	*	-	-	-	-
TCMSV	Community Midwifery Visit	Service Code	Visits	*	-	-	-	-	-
TVMed	Community Medical Services Vaccinations	Service Code	Vaccinations	*	-	-	-	-	-
TCMed	Community Medical Services Client	Service Code	Clients Seen	*	-	-	-	-	-
TTHPYS	Community Therapy Services Client	Service Code	Clients Seen	*	* [a]	* [c]	-	-	-
TOCS	Community Services Other Attendance	Service Code	Attendances	*	*	* [d]	-	-	-
TMHi	Mental Health Services Inpatient	Service Code	Occupied Bed Days	*	*	*	-	-	-
TMHii	Mental Health Services Booked Appointment	Service Code	First Appointments	*	*	*	-	-	-
TMHiii	Mental Health Services Booked Appointment	Service Code	Follow up Appointments	*	*	*	-	-	-
TMHiv	Mental Health Services Domiciliary	Service Code	Domiciliary Visits	*	*	*	-	-	-
TMHv	Mental Health Services Secure Unit	Service Code	Occupied Bed Days	*	*	-	-	-	-
TMHvi	Mental Health Services Specialist Services Inpatient	Service Code	Occupied Bed Days	*	*	-	-	-	-
TMHvii	Mental Health Services Specialist Services Booked Appointments	Service Code	First Appointments	*	*	-	-	-	-
TMHviii	Mental Health Services Specialist Services Booked Appointments	Service Code	Follow up Appointments	*	*	-	-	-	-

TMHix	Mental Health Services Specialist Teams	Service Code	Clients Seen	*	*	-	-	-	-
TMHx	Mental Health Services Day Care Facilities	Service Code	Patient Days	*	-	-	-	-	-
Tambresi	Paramedic Services provided by Ambulance NHS Trust Category A	Code	Responses	*	-	-	-	-	-
Tambresii	Paramedic Services provided by Ambulance NHS Trust Category B/C	Code	Responses	*	-	-	-	-	-
Tambresiii	Paramedic Services provided by Ambulance NHS Trust Emergency Transfers	Code	Responses	*	-	-	-	-	-
Tambresiv	Paramedic Services Provided by Ambulance NHS Trusts Other Services	Code	Responses	*	-	-	-	-	-
Tawmbresv	Paramedic Services Provided by Ambulance NHS Trusts Urgent	Code	Responses	*	-	-	-	-	-
Tambinci	Paramedic Services Provided by Ambulance NHS Trusts Category A Incident	Code	Incidents	*	-	-	-	-	-
Tambincii	Paramedic Services provided by Ambulance NHS Trust Category B/C Incident	Code	Incidents	*	-	-	-	-	-
Tambinciii	Paramedic Services provided by Ambulance NHS Trust Emergency Transfer Incident	Code	Incidents	*	-	-	-	-	-
Tambinciv	Paramedic Services Provided by Ambulance NHS Trusts Other Services Incident	Service Code	Incidents	*	-	-	-	-	-
Tambincv	Urgent Incident	Code	Incidents	*	-	-	-	-	-

Notes to NHS Trusts Reference Costs

* = data available.
\- = data not available.
(a) = 2002 TTHPYS includes both 2003 TTHPYS (2) and TTHPYS data. However, data in 2003 are differentiated between ADULT and CHILD patients, whereas the one for 2002 are not. Further, 2003 data are registered separately for Assessment and Treatment. This distinction is not made in 2002 data.
(b) = TCNS is not comparable to the 2003 TCNS, but it is to the 2003 TCNS (2). However, data in 2003 are differentiated between ADULT and CHILD patient, whereas in both 2002 and 2001 they are not.
(c) = 2001 TTHPYS data for each therapy are organised by how many times a client has been seen. No distinction is made between ADULT and CHILD patient and between Assessment and Treatment.
(d) = Includes 1 more variable: Needle Exchange Schemes.

Organisation	Acronym	Activity Category	Type of Activity Code	Units of Activity	2003	2002	2001	2000	1999	1998
PC TRUSTS	**PCTELIP**	Elective inpatients	HRG	FCE	*	*	*	-	-	-
	PCTNELIP	Non-Elective inpatients	HRG	FCE	*	*	*	-	-	-
	PCTDC	Day cases HRG	HRG	FCE	*	*	*	-	-	-
	PCTCCS	Critical Care Services	Service Code	Bed Days	*	-	-	-	-	-
	PCTOP HRG	Outpatient HRG	HRG	Attendances	*	*	*	-	-	-
	PCTOPS F	Outpatient First Attendance	Speciality Code	First Attendances	*	*	*	-	-	-
	PCTOPS FU	Outpatient Follow up Attendance	Speciality Code	Follow up Attendances	*	*	*	-	-	-
	PCTOPS MAT	Outpatient Maternity Services Attendance	Service Code	Various	*	*	-	-	-	-
	PCTOPWA	Outpatient Ward Attenders	Service Code	Ward Attenders	*	-	-	-	-	-
	PCTCDCWA	Ward Attenders	Service Code	Ward Attenders	*	-	-	-	-	-
	PCTRDNA	Regular Day and Night Admissions	Service Code	Admissions	*	-	-	-	-	-
	PCTAH	Audiological Services	Service Code	Various	*	*	-	-	-	-
	PCTPATH	Direct Access: Pathology Services Test	Service Code	Tests	*	*	*	-	-	-
	PCTRADIO	Direct Access: Radiology Services Test	Service Code	Tests	*	*	*	-	-	-
	PCTCMTESTS	Direct Access: Clinical Measurement	Service Code	Tests	*	-	-	-	-	-
	PCTTPHYS (2)	Direct Access Therapy Services	Service Code	Clients Seen	*	* [a]	-	-	-	-
	PCTA&E	Accident and Emergency Minor Injury Unit Services	Service Code	Attendances	* [t]	* [b]	-	-	-	-
	PCTRENALD	Renal Dialysis Sessions	Service Code	Sessions	*	-	-	-	-	-
	PCTREHAB	Rehabilitation Services Occupied Bed Days	Service Code	Occupied Bed Days	*	*	-	-	-	-
	PCTDCF	Day Care Facilities Regular Attendance	Service Code	Patient Days	*	-	-	-	-	-
	PCTHAH	Hospital at Home Team Client	Service Code	Team Clients Seen	*	-	-	-	-	-
	PCTCNS (2)	Community Nursing Services	Service Code	Various	*	-	-	-	-	-
	PCTCNS	Community/Outreach Specialist Nursing Services	Service Code	Clients Seen	*	* [c]	* [c]	-	-	-
	PCTCMS	Community Midwifery Services HRG	HRG	Numbers	*	*	-	-	-	-

PC TRUSTS	**PCTCMSV**	Community Midwifery Visit	Service Code	Visits	*	-	-	-	-	-
	PCTVMed	Community Medical Services Vaccinations	Service Code	Vaccinations	*	-	-	-	-	-
	PCTCMed	Community Medical Services Client	Service Code	Clients Seen	*	-	-	-	-	-
	PCTTHPYS	Community-based Therapy Services Client	Service Code	Clients Seen	*	* (a)	* (d)	-	-	-
	PCTOCS	Community Services Other Attendance	Service Code	Attendances	*	*	* (e)	-	-	-
	PCTMHi	Mental Health Services Inpatient Services	Service Code	Occupied Bed Days	*	*	*	-	-	-
	PCTMHii	Mental Health Services Booked Appointment	Service Code	First Appointments	*	*	*	-	-	-
	PCTMHiii	Mental Health Services Booked Appointment	Service Code	Follow up Appointments	*	*	*	-	-	-
	PCTMHiv	Mental Health Services Domiciliary Visit	Service Code	Domiciliary Visits	*	*	*	-	-	-
	PCTMHv	Mental Health Services Secure Unit	Service Code	Occupied Bed Days	*	*	-	-	-	-
	PCTMHvi	Mental Health Services Specialist Services Inpatient	Service Code	Occupied Bed Days	*	-	-	-	-	-
	PCTMHvii	Mental Health Services Specialist Services Booked Appointment	Service Code	First Appointments	*	-	-	-	-	-
	PCTMHviii	Mental Health Services Specialist Services Booked Appointment	Service Code	Follow up Appointments	*	-	-	-	-	-
	PCTMHix	Mental Health Services Specialist Teams	Service Code	Clients Seen	*	*	-	-	-	-
	PCTMHx	Mental Health Services Day Care Facilities	Service Code	Patient Days	*	-	-	-	-	-

Notes to PC Trusts Reference Costs

*	=	data available.
-	=	data not available.
t	=	There is a discrepancy between the hardcopy and file version for the data for this year. Hardcopy classifies PCTA&E as service code, whereas in the file it is classified as HRG.
(a)	=	2002 PCTTHPYS includes both 2003 PCTTHPYS (2) and PCTTHPYS data. However, data in 2003 are differentiated between ADULT and CHILD patients, whereas the one for 2002 are not. Further, 2003 data are registered separately for Assessment and Treatment. This distinction is not made in 2002 data.
(b)		PCTA&E is not comparable to that in 2003.
(c)	=	PCTCNS is not comparable to the 2003 PCTCNS, but it is to the 2003 PCTCNS (2). However, data in 2003 are differentiated between ADULT and CHILD patient, whereas in both 2002 and 2001 they are not.
(d)	=	2001 PCTTHPYS data for each therapy are organised by how many times a client has been seen. No distinction is made between ADULT and CHILD patient and between Assessment and Treatment.
(e)	=	Includes 1 more variable: Needle Exchange Schemes.

Organisation	Acronym	Activity Category	Type of Activity Code	Units of Activity	2003	2002	2001	2000	1999	1998
Personal Medical Services and Pilot	**pmsDC**	Day Cases	HRG	FCE	*	*	*	-	-	-
	pmsOP HRG	Outpatient HRG	HRG	Attendances	*	*	*	-	-	-
	pmsOPS F	Outpatient First Attendance	Speciality Code	First Attendances	*	*	*	-	-	-
	pmsOPS FU	Outpatient Follow up Attendance	Speciality Code	Follow up Attendances	*	*	*	-	-	-
	pmsAH	Audiological Services	Service Code	Various	*	-	-	-	-	-
	pmsPATH	Direct Access Pathology Services Test	Service Code	Tests	*	* (a)	* (a)	-	-	-
	pmsCMTESTS	Direct Access Clinical Measurement Data	Service Code	Tests	*	-	-	-	-	-
	pmsCNS	Community Nursing Services	Service Code	Various	*	-	-	-	-	-
	pmsCNS (2)	Community/Outreach Specialist Nursing Services	Speciality Code	Clients seen	*	* (b)	* (b)	-	-	-
	pmsCMed	Community Medical Services Client Data	Speciality Code	Clients seen	*	-	-	-	-	-
	pmsTHPYS	Community-based Therapy Services Client Data	Speciality Code	Clients seen	*	* (c)	* (d)	-	-	-
	pmsOCS	Community Services Other Attendance Data	Speciality Code	Clients seen	*	*	*	-	-	-
	pmsMHii	Mental Health Services Booked Appointment Data	Service Code	First Appointments	*	*	*	-	-	-
	pmsMHiii	Mental Health Services Booked Appointment Data	Service Code	Follow up Appointments	*	*	-	-	-	-

Notes to Personal Medical Services and Pilot Reference Costs

* = data available.
\- = data not available.
(a) = Not comparable to 2003 data.
(b) = pmsCNS is not comparable to the 2003 pmsCNS, but there are to some extent comparable to 2003 pmsCNS (2).
(c) = 2002 pmsTHPYS are not directly comparable to 2003 data. 2003 data are differentiated between ADULT and CHILD patients, whereas the one for 2002 are not. Further, 2003 data are registered separately for Assessment and Treatment. This distinction is not made in 2002 data.
(c) = 2001 pmsTHPYS are not directly comparable to 2003 data. Data for each therapy are organised by how many times a client has been seen. No distinction is made between ADULT and CHILD patient and between Assessment and Treatment.

Trust Financial Returns (TFR2, 3, 6) data description and data availability

Trust Financial Returns cover the activity and financial data returns for all NHS Trusts and are listed in the second column. These are available for almost all of the types of services or specialties or types of staff or types of expenditure listed in the key. A few minor changes have occurred over time in for instance the categories of expenditure collected over time.

Return	Description of data return	Type of data		Key	Data return available for each of the following	1996/97	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03
TFR2A	Maincode: 056 Patient Days	Patients Using a Bed (including day cases)	Medical specialties	101	Paediatrics	y	y	y	y	y	y	y
	Maincode: 057 Finished Consultant/GP Episodes	Patients Using a Bed (including day cases)		102	Geriatrics	y	y	y	y	y	y	y
	Maincode: 058 Net Expenditure (£)	Patients Using a Bed (including day cases)		103	Cardiology	y	y	y	y	y	y	y
	Maincode: 059 Cost per patient day [058/056] - (£)	Patients Using a Bed (including day cases)		104	Dermatology	y	y	y	y	y	y	y
	Maincode: 060 Cost per episode [058/057] - (£)	Patients Using a Bed (including day cases)		105	Infectious Diseases	y	y	y	y	y	y	y
	Maincode: 061 First Attendances	Outpatients (including nurse clinic and ward attenders)		106	Medical Oncology	y	y	y	y	y	y	y
	Maincode: 062 Total Attendances	Outpatients (including nurse clinic and ward attenders)		107	Neurology	y	y	y	y	y	y	y
	Maincode: 063 Net Expenditure (£)	Outpatients (including nurse clinic and ward attenders)		108	Rheumatology	y	y	y	y	y	y	y
	Maincode: 064 Cost per attendance [063/062] - (£)	Outpatients (including nurse clinic and ward attenders)		109	Gastroenterology	y	y	y	y	y	y	y
				110	Haematology							
				111	Clinical Immunology & Allergy							
				112	Thoracic Medicine							
				113	Genito-urinary Medicine							
				114	Nephrology							
				115	Rehabilitation Medicine							
				125	Other Medicine							
TFR2B	Maincode: 056 Patient Days	Patients Using a Bed (including day cases)	Surgical specialties	200	General Surgery	y	y	y	y	y	y	y
	Maincode: 057 Finished Consultant/GP Episodes	Patients Using a Bed (including day cases)		201	Urology	y	y	y	y	y	y	y
	Maincode: 058 Net Expenditure (£)	Patients Using a Bed (including day cases)		202	Orthopaedics	y	y	y	y	y	y	y

	Maincode: 059 Cost per patient day [058/056] - (£)	Patients Using a Bed (including day cases)		203	ENT	y	y	y	y	y	y	y
	Maincode: 060 Cost per episode [058/057] - (£)	Patients Using a Bed (including day cases)		204	Ophthalmology	y	y	y	y	y	y	y
	Maincode: 061 First Attendances	Outpatients (including nurse clinic and ward attenders)		205	Gynaecology	y	y	y	y	y	y	y
	Maincode: 062 Total Attendances	Outpatients (including nurse clinic and ward attenders)		206	Dental Specialties	y	y	y	y	y	y	y
	Maincode: 063 Net Expenditure (£)	Outpatients (including nurse clinic and ward attenders)		207	Neuro-surgery	y	y	y	y	y	y	y
	Maincode: 064 Cost per attendance [063/062] - (£)	Outpatients (including nurse clinic and ward attenders)		208	Plastic Surgery	y	y	y	y	y	y	y
				209	Cardiothoracic							
				210	Paediatric Surgery							
				301	Obstetrics							
				302	General Practice							
TFR2C	Maincode: 056 Patient Days	Patients Using a Bed (including day cases)	Psychiatric specialties	401	Learning Difficulties/Disabilities	y	y	y	y	y	y	y
	Maincode: 057 Finished Consultant/GP Episodes	Patients Using a Bed (including day cases)		402	Mental Illness	y	y	y	y	y	y	y
	Maincode: 058 Net Expenditure (£)	Patients Using a Bed (including day cases)		403	Child & Adolescent Psychiatry	y	y	y	y	y	y	y
	Maincode: 059 Cost per patient day [058/056] - (£)	Patients Using a Bed (including day cases)		404	Forensic Psychiatry	y	y	y	y	y	y	y
	Maincode: 060 Cost per episode [058/057] - (£)	Patients Using a Bed (including day cases)		405	Psychotherapy	y	y	y	y	y	y	y
	Maincode: 061 First Attendances	Outpatients (including nurse clinic and ward attenders)		406	Old Age Psychiatry	y	y	y	y	y	y	y
	Maincode: 062 Total Attendances	Outpatients (including nurse clinic and ward attenders)	Other specialties	501	General Practice (excluding maternity)	y	y	y	y	y	y	y
	Maincode: 063 Net Expenditure (£)	Outpatients (including nurse clinic and ward attenders)		502	Radiotherapy	y	y	y	y	y	y	y
	Maincode: 064 Cost per attendance [063/062] - (£)	Outpatients (including nurse clinic and ward attenders)		503	Pathological Specialties & Radiology	y	y	y	y	y	y	y
				504	Anaesthetics							
				505	Accident and Emergency							
TFR2D	Maincode: 056 Patient Days	Patients Using a Bed (including day cases)	Supra-district services	601	Renal Dialysis (inc. CAPD)	y	y	y	y	y	y	y
	Maincode: 057 Finished Consultant/GP Episodes	Patients Using a Bed (including day cases)		602	Renal Transplant	y	y	y	y	y	y	y

Maincode: 058 Net Expenditure (£)	Patients Using a Bed (including day cases)		603 Open Heart Surgery	y	y	y	y	y	y	y
Maincode: 059 Cost per patient day [058/056] - (£)	Patients Using a Bed (including day cases)	Supra-regional services	701 Spinal Injuries Services	y	y	y	y	y	y	y
Maincode: 060 Cost per episode [058/057] - (£)	Patients Using a Bed (including day cases)		702 Cranio Facial Services	y	y	y	y	y	y	y
Maincode: 061 First Attendances	Outpatients (including nurse clinic and ward attenders)		703 Neonatal & Infant Cardiac Surgery	y	y	y	y	y	y	y
Maincode: 062 Total Attendances	Outpatients (including nurse clinic and ward attenders)		704 Chorioncarcinoma	y	y	y	y	y	y	y
Maincode: 063 Net Expenditure (£)	Outpatients (including nurse clinic and ward attenders)		705 National Poisons Information Service	y	y	y	y	y	y	y
Maincode: 064 Cost per attendance [063/062] - (£)	Outpatients (including nurse clinic and ward attenders)		706 Liver Transplantation	y	y	y	y	y	y	y
			707 Specialist Paediatric Liver Services							
			708 Heart Transplantation							
			709 Endoprosthetic Services for Primary Bone Tumours							
			710 Psychiatric Services for Deaf People							
			711 Retinoblastoma Services							
			712 Steriotactic Radiosurgery							
			713 Severe Combined Immunodeficiency and Related Disorders (SCID)							
			714 Extra Corporeal Membrane Oxygenation (ECMO)							
			715 Ocular Oncology							
			716 Diagnosis and Management of Gauchers Disease							
			717 Severe Personality Disorder Service							
			718 Inpatient Psychiatric Service for Deaf Cildren and Adolescents							
			719 Treatment of Established Intestinal Failure							

				720 Gynaecological Reconstruction								
				721 Small Bowel Transplantation								
				722 Total Anorectal Reconstruction								
				723 Amyloidosis								
				724 Pseudomyxoma Peritonei								
				725 Pulmonary Thrombo Endarectomy								
				726 Prion Disease								
				727 Paediatric Bladder Exstrophy								
				728 Rare Neuromuscular Diseases								
TFR2E	Maincode: 065 Expenditure (£)	Patients using accident & emergency services			y	y	y	y	y	y	y	
	Maincode: 066 Attendances	Patients using accident & emergency services			y	y	y	y	y	y	y	
	Maincode: 067 New Cases	Patients using accident & emergency services			y	y	y	y	y	y	y	
	Maincode: 068 First Attendances	Day Care Patients	General Medicine	901 Younger physically disabled	y	y	y	y	y	y	y	
	Maincode: 069 Total Attendances	Day Care Patients		902 General	y	y	y	y	y	y	y	
	Maincode: 070 Expenditure (£)	Day Care Patients		903 Neurology	y	y	y	y	y	y	y	
	Maincode: 071 Cost per attendance [070/069] - (£)	Day Care Patients		904 Rheumatology	y	y	y	y	y	y	y	
				905 Geriatric (elderly care)								
				906 Learning disabilities								
			Mental Illness	907 Alcoholism								
				908 Drug abuse								
				909 Psychogeriatric								
				910 General								
				911 Child & Adolescent Psychiatry								
				912 Old Age Psychiatry (elderly mentally ill)								
				913 Paediatrics								
			Other Specialties	914 Alcoholism								
				915 Drug abuse								
				916 Younger physically disabled								
				917 Psychogeriatric								
				918 General								

TFR2F	Maincode: 072 Number of Work Units	Health Programme Analysis	Paediatric Community Services	101	Assessment & development	y	y	y	y	y	y	y
	Maincode: 073 Expenditure (£)	Health Programme Analysis		102	Vaccination and immunisation	y	y	y	y	y	y	y
	Maincode: 074 Unit Costs	Health Programme Analysis		103	Dental	y	y	y	y	y	y	y
				104	Other professional advice & support							
				105	Mental Illness							
				106	Learning disabilities							
			General	107	Professional advice & support							
				108	District Nursing							
				109	Specialist Nursing							
				110	Children/ School Nursing							
				111	Health Visiting							
				112	Chiropody/ Podiatry							
				113	Other Paramedical							
				114	Screening							
				115	Home dialysis							
				116	Family planning							
				117	Speech Therapy							
				118	Occupational Therapy							
				119	Physiotherapy							
				120	Dietrtrics							
				121	Community Dentistry							
				122	Maternity Services							
				123	Services to GPs under open access							
				124	Services for Local Authorities							
				125	Health Education and Promotion							
				126	Other Services							
TFR3A	NHS Staff Salaries and Wages - Expenditure (£)		Medical staff	5	Total Senior Managers & Managers (inc. executive directors)	y	y	y	y	y	y	y
				7	SHMOs Medical Assistants							
				8	Assorted Specialists							

			9	Staff Grade Practitioners
			10	Senior Registrars
			11	Registrars
			12	Senior House Officers
			13	House Officers
			14	Hospital Practitioners
			15	Clinical Assistants/ BTS Sessions
			16	Staff Fund Payments
			17	Senior Clinical Medical Officers
			18	Clinical Medical Officers
			19	Consultants in Public Health Medicine
			20	Consultants
			21	Other Career Grades
			22	All Registrars and Senior Registrars
			23	SHOs and HOs
			24	Other Medical Hospital Grades
			25	Total medical staff (including locums)
		Dental staff	27	SHDOs Assistant Dental Surgeons
			28	Assistant Specialists
			29	Staff Grade Practitioners
			30	Senior Registrars
			31	Registrars
			32	Senior Dental House Officers
			33	Dental House Officers
			34	Dental Practitioners
			35	Consultants
			36	Other Career Grades
			37	All Registrars and Senior Registrars
			38	SDHOs and DHOs

				39 Other Dental Hospital Grades	
				40 Total dental staff (including locums)	
			Nursing, midwifery & health visiting staff	42 Other Nursing, Midwifery & Health Visiting Staff	
				49 Nursing, Midwifery & Health Visiting Learners	
				56 Nurse Consultants	
				57 Nurse Managers	
				58 Qualified	
				59 Unqualified	
				60 Total Nursing Midwifery & Health Visiting Staff	
			Scientific, therapeutic & technical staff	61 Professions Allied to Medicine (excludes speech therapists)	
				70 Scientific and Professional Staff (includes speech therapists)	
				72 Professional and Technical -. S*(R to T)	
				74 Other Scientific, Therapeutic & Technical Staff	
				75 Total Scientific, Therapeutic & Technical Staff	
				110 Administrative & clerical	
				120 Healthcare assistants & other support staff (to include ancilliary)	
				130 Maintenance & works staff	
				140 Ambulance staff	
				145 Other employees	
				150 Total NHS staff salaries & wages	

TFR3B	Non NHS staff (Agency, etc.) (by NHS classification) Salaries and Wages - Expenditure (£)			151 Medical	y	y	y	y	y	y	y
				152 Dental							
				153 Nursing, midwifery and health visiting							
				154 Scientific, Therapeutic & Technical Staff							
				160 Administrative & Clerical							
				161 Healthcare Assistants & Other Support Staff							
				162 Maintenance & Works Staff							
				163 Ambulance Staff							
				164 Other Employees							
				170 Total Non NHS staff salaries & wages							
				180 Chairman & Non-Executive Directors Remuneration							
				190 Total Medical & Dental London Weighting (inc. above)							
				200 Total Revenue Expenditure on salaries & wages (S/C150 + 170 + 180)							
TFR3C	Non Pay Expenditure (£)		Clinical supplies & services	301 Occupational/ Industrial Therapy Materials & Equipment	y	y	y	y	y	y	y
				302 Drugs							
				303 Medical Gases							
				304 Dressings							
				305 Medical & Surgical Equipment - Purchase							
				306 Medical & Surgical Equipment - Maintenance							
				307 X-Ray Equipment - Purchase							
				308 X-Ray Film & Chemicals - Purchase							

			309	X-Ray Equipment - Maintenance
			310	Appliances
			311	Artificial Limb & Wheelchair Hardware
			312	Laboratory Equipment - Purchase
			313	Laboratory Equipment - Maintenance
			314	Other Clinical Supplies
			320	Total Clinical Supplies and Services
		General supplies & services	321	Provisions & Kitchen
			322	Contract Hotel Services (incl. Cleaning)
			324	Uniforms & Clothing
			325	Laundry & Cleaning Equipment
			326	Laundry: External Contracts
			327	Hardware & Crockery
			328	Bedding & Linen: Disposable
			329	Bedding & Linen: Non - disposable
			330	Total General Supplies and Services
		Establishment expenditure	331	Printing & Stationery
			332	Postage
			333	Telephone - rental
			334	Telephone - Other (including calls)
			335	Advertising
			336	Travel, Subsistence & Removal Expenses
			337	Removal Expenses

				338	Other transport costs (includes transport & moveable plant)								
				340	Total Establishment Expenditure								
TFR3D	Non Pay Expenditure (£)		Premise and fixed plant	341	Fuel & Oil	y	y	y	y	y	y	y	
				342	Maintenance - Equipment & Materials								
				343	Maintenance - External Contracts								
				344	Hire of Transport								
				345	Hospital Car Service								
				346	Miscellaneous Transport Expenses (inc. Vehicle Insurance)								
				351	Coal								
				352	Oil								
				353	Electricity								
				354	Gas								
				355	Other Fuels (including oil and coal)								
				356	Water and Sewerage								
				357	Cleaning Equipment/ Materials								
				358	External General Services Contracts								
				359	Furniture, Office & Computer Equipment								
				360	Office Equipment								
				361	Computer Hardware & Software: Purchases (inc.rent & leasing)								
				362	Computer Hardware,Maintenance Contracts & Data Processing Contracts								
				363	External Contracts for Data Processing								
				364	Business Rates								

			365	Rent
			366	Building and Engineering Equipment
			367	Building & Engineering Contracts
			368	Building Maintenance: Equipment/ Materials
			369	Building Maintenance: External Contracts
			370	Gardening & Farming: Equipment/ Materials
			371	Gardening & Farming: External Contracts
			380	Total Premises and Fixed Plant
		Depreciation	381	Depreciation on owned assets (Capital charges)
			383	Depreciation on donated assets
			384	Total Depreciation
			385	Fixed Asset Impairments
		Miscellaneous	386	Voluntary Sector
			387	Commercial Sector
			388	Total purchase of healthcare from non-NHS bodies
			390	Total external contract staffing and consultancy services
			391	Srudents' Bursaries
			392	Patients Allowances
			393	Auditors Remuneration
			394	Gross Redundancy Payments
			395	Net Bank Charges
			396	Patients' Travelling Expenses
			397	Other Insurance Charges (not property or vehicles)
			398	Other Miscellaneous
			400	Total Miscellaneous

				500 Total non-pay revenue expenditure								
TFR3E	Analysis of Expenditure by Type (£)			600 Total Salaries & Wages (Subcode 200) 610 Total Non-Pay Revenue Expenditure (Subcode 500) 630 Services from other NHS bodies (not recharges) - Non-Healthcare 635 Services from other NHS bodies (not recharges) - Subcontracted Healthcare 640 Total revenue expenditure	y	y	y	y	y	y	y
TFR6	Ambulance Services Expenditure (£)			101 Emergency patient transport services 102 Non-emergency patient transport services 103 All other expenditure 110 Total expenditure	y	y	y	y	y	y	y

CIPFA data

In addition to the TFR data listed above the CIPFA data also includes Trust Accounting Cost (TAC) data for all NHS Trusts. This is available for the same years as the TFR data.

CIPFA data also covers PFR and ASF data for all PCTs similar to the TFR and TAC data for NHS Trusts, as well as HFR and HAA data for all Health Authorities and Strategic Health Authorities (through time).

Healthcare Commission performance data and data availability

Data are available for acute NHS Trusts for the three years 2000/01 to 2002/03. Specialist Trusts, Mental Health Trusts and Ambulance Trusts started receiving star ratings in 2001/02 and two years of data exist. PCTs started receiving ratings in 2002/03. The next rating is due in summer 2004 and will cover all organisations. The performance indicators covered by this data source will be listed in due course.

Hospital activity statistics

This is available for the following years: 1996/97 to 2001/02, at the following website:
http://www.performance.doh.gov.uk/hospitalactivity/index.htm
and covers beds open overnight, critical care beds, residential care beds, A&E attendances, admissions from A&E, total time in A&E, outpatient attendances, waiting in outpatients, ward attendances, cancelled operations, NHS written complaints, imaging and radiodiagnostics, supporting facilities and day care attendances. Details of this data will be listed in due course.

GP activity

GPs diagnose patients, they prescribe treatment, they attempt to prevent ill health by giving lifestyle advice (primary prevention), they screen patients to detect disease earlier (secondary prevention), they refer patients who they either cannot diagnose or

cannot treat, and they provide after care for patients treated in the secondary sector. With detailed information on the content of consultations it should be possible to distinguish types of GP activity but quantifying their effects on health sector output (either in terms of QALY gains or otherwise) will be a non-trivial task.

There are no routine administrative data on the number of consultations, let alone their content. A small amount of general practice activity has been covered by routinely collected series on activities such as cervical screening and vaccination and immunisation which are remunerated as part of the old General Medical Services (GMS) contract. There is less centrally collected data on the activity of the 35% of general practices which hold Primary Medical Service (PMS) contracts since these are negotiated and administered locally at PCT level. The new GMS contract for GPs contains a large number of quality related performance measures for general practice and hence may be a useful source in future if the data generated are centrally collected at a suitable level of disaggregation. But these data will not cover the practices on PMS contracts. Other possible sources to derive estimates of activity include: current national population surveys (such as the General Household Survey, the British Household Panel Survey, the Health Survey for England, DH patient surveys) and practice level data sets (such as the General Practice Research Database).

Other data sources

Other data sources still to be explored include the full Körner activity data, sources of GP activity data, GPRD from the Medicines Control Agency, the General Household Survey and the British Household Panel Survey for utilisation data, Workforce Survey data and sources for dental care, community care, and social care data.