Getting out what we put in: how productive is the NHS in England?

Adriana Castelli, Mauro Laudicella, Andrew Street[1] and Padraic Ward
Centre for Health Economics, University of York
[1]Corresponding author: ads6@york.ac.uk

Not to be quoted without authors' permission

Paper presented at the Health Economist' Study Group meeting in Sheffield, 22-24 July 2009

Abstract:
The English NHS has received significant increases in funding over the last few years. We assess what this funding has achieved. We construct an index of output growth, incorporating all care provided to NHS patients. Quality is captured by survival rates, waiting times and disease management. We measure growth in labour, intermediate and capital inputs. Productivity is assessed by comparing output growth with growth in inputs. We analyse data from sources including the hospital episode statistics, reference cost returns, workforce census, financial returns and prescription pricing authority. Between 1998/9 and 2003/4 output growth lagged behind growth in inputs, driven by pay awards, the European Working Time Directive and investment in equipment and buildings. Since 2004/5, NHS output has increased at a faster rate and input growth has slowed, particularly as less use is made of agency staff. Consequently recent NHS productivity growth has been constant, at least.

Word count: 6447

1. Introduction

The national accounts measure aggregate economic activity, summarised using indicators of economic performance such as gross domestic product, national income and inflation. First published by the United States in 1947, other countries soon started producing their own accounts, with the United Nations setting out international conventions. But it is only recently that the contribution of the health system has been properly recognised in the national accounts, mainly because of difficulties in defining and measuring the 'output' of health care.

It is important to measure health care output accurately so that proper account can be taken of the resources devoted to health care. The English NHS has received significant increases in funding over the last few years and we assess what this funding has achieved. We construct a comprehensive index of output growth, incorporating all care provided to NHS patients and we implement a means of dealing with changes to how health services are defined. Improvements in quality are captured by survival rates, waiting times and disease management. Insofar as is possible our index of input growth captures all inputs into health service production growth and we improve on how capital has been measured in the past.

2. Methods

In static terms, productivity can be defined as a simple ratio that compares the volume of output produced to the volume of input utilised in the production process at a particular point of time. Of more interest is the change in productivity over time: if output growth exceeds input growth, this is interpreted as an improvement in productivity. Total factor productivity growth is calculated by dividing an index of output growth by an index of input growth:

$$\Delta TFP = \left[I/Z \right] - 1 \qquad (1)$$

Where ΔTFP is total factor productivity growth, I is the index of output growth and Z is the index of input growth. In order to estimate total factor productivity, it is therefore necessary to correctly define and measure its underlying components. Section 2.1 address the methodological challenges involved in measuring NHS outputs, whilst section 2.2 addresses the issues involved in dealing with NHS inputs.

2.1 Indices of output growth

Eurostat (2001) defines healthcare output as "the quantity of health care received by patients, adjusted to allow for the qualities of services provided, for each type of health care. The quality of health care received by patients should be measured in terms of complete treatments".

There are two major challenges in meeting this definition for the purposes of constructing an index of output growth. First, it is necessary to quantify the volume of health care accurately. As patients have varied health care requirements and receive very different packages of care, this necessitates classifying patients into reasonably homogenous 'output' groupings. Second, in order to aggregate these output groups into a single index, some means of assessing their relative value is required. We discuss these issues in turn.

Quantifying the volume of health care output in terms of 'completed treatments' is difficult for two main reasons. First, many patients receive a range of interventions from different providers, in a variety of settings. Most countries, including England, lack the informational capability to track patients across different settings. Consequently we cannot capture accurately the full treatment pathway. Second, it is not always straightforward to determine when treatment has been completed. Indeed, for patients with chronic or terminal conditions who require care over a long period of time, treatment may not be considered 'complete' until the patient has died. Rather than quantifying 'complete treatments', it is common practice to define output in the health sector by counting the amount of each type of output that is undertaken in each health care setting. An output might be a consultation with a general practitioner (GP), an angioplasty involving a stay in hospital or a visit to the outpatient department. We define x_j as the number of patients who have output type *j*, where *j*=1…*J*.

In order to construct a measure of total output it is necessary to attach a relative value to each type of output (v_j), allowing diverse activities to be aggregated. If values were available, a Laspeyres output index would take the following form[1]:

$$I^v = \frac{\sum_{j=1}^{J} x_{jt+1} v_{jt}}{\sum_{j=1}^{J} x_{jt} v_{jt}} \qquad (2)$$

[1] We specify indices in Laspeyres form throughout. It is straightforward to consider alternative formulations.

The practical problem in constructing this index, particularly for non-market sectors, is that there is no readily available means of determining the relative value of different activities. As an alternative, the convention in the National Account has been to use costs to reflect the value of non-market outputs. Hence, if cost weights are used instead of value weights to adjust NHS outputs, equation (2) becomes:

$$I^c = \frac{\sum_{j=1}^{J} x_{jt+1} c_{jt}}{\sum_{j=1}^{J} x_{jt} c_{jt}} \qquad \textbf{(3)}$$

Where c_{jt} indicates the cost of output *j* at time *t*. We now consider two issues that must be addressed if this index is to provide an accurate reflection of output growth in the NHS. The first concerns the stability of output categorisation over time. We discuss this in section 2.1.1. Second we consider whether changes in value over time, driven primarily by changes in the quality of health care activities, are adequately reflected by changes in cost. In section 2.1.2 we describe how equation (3) can be extended in order to capture changes in quality.

2.1.1 Changes in output categories

The way that output categories are defined need not stay constant over time. One reason for this is that new technologies appear, as they do in all sectors of the economy. But of more consequence in the health sector is that the classification systems used to describe output categories are often subject to substantial revision. This makes it difficult to make direct comparisons of output from one period to the next and, therefore, to calculate growth rates.

Traditional methods to calculate output growth require output categories to be consistent across adjacent time periods (Eurostat/Commission of the European Communities, et al. 1993). However, categorisation of health service output in England has been subject to wholesale revision in both the Reference Cost collections and in the Hospital Episode Statistics in recent years. In order to construct an output index, some means of overcoming the requirement for consistent definition of output categories over time is required.

There are three ways to deal with the problem of categorisation changes. Method A is the *traditional* approach, and entails inclusion of output categories only if information is available in two successive years. An example is the Consumer Price Index, which it would be impractical to calculate using information about all goods and services traded in the economy. Instead, the CPI is based on price changes for a basket of some 650 items, the composition of this basket remaining constant (Pike R. 2008). While this avoids re-categorisation problems, the assumption is that the basket is representative of all traded items. Rather than being based on a selected sub-set of activities, the NHS output index aims to be comprehensive.

Method B involves *mapping* of new and retiring activities. This strategy was adopted in Dawson et al. (2005) and is employed by ONS (Office for National Statistics 2009). This requires that all new and retiring categories can be mapped to each another in some way. This cannot be guaranteed. The mapping currently employed by ONS means that their index of health output captures only 70% of healthcare expenditure (Office for National Statistics 2009).

The fundamental problem in calculating a Laspeyres index is that when a new output category (x_j^N) is introduced in $t+1$, there is no value for its cost in the previous (base) year c_{jt}.[2] Method C involves *imputing* values where cost data are missing for any particular time period, which requires a judgment to be made about the nature of the (unobservable) price for the missing period. The US Bureau of Labor Statistics assumes that, for a new product, the relationship between its earlier and current price reflects the change observed for other goods in the relevant sector (Cutler, et al. 2001). Similarly the approach adopted here is to base the unobservable value for c_{jt} on the observed value c_{jt+1}. The use of c_{jt+1} necessitates an assumption about how costs would have changed for the specific output type from the previous period, had these outputs been provided. This involves specifying a deflator, γ_j, so that $c_{jt} = \gamma_j \, c_{jt+1}$ can be calculated.

To illustrate the implications of adopting one method over another we consider ten output categories that are subject to different volumes over time. In order to assess the pure volume effect of the alternative methods, we assume a common set of cost weights (ie each unit of output is of equal value) and no inflation. Table 1 provides illustrative data for ten output categories covering the spectrum of cases that present themselves when dealing with the construction of output growth indices, these being:

- Categories subject to no categorisation changes, which present no problems. Categories A and B are examples of categories that are recorded throughout the whole time period.
- Introduction of a new category and subsequent retirement. Category C is introduced at time t_1 and retired at t_3.
- Retirement and subsequent re-introduction of an original category. Category D is retired in t_1, and then re-introduced at t_3.
- Subsequent disaggregation of an original category. The activities recorded under category E at t_0 and t_1 are disaggregated at time t_2 to form two categories F and G.
- Subsequent aggregation of originally separate categories. Output in categories H and I was itemised separately until t_1 but was amalgamated into the single Category J at time t_2.

The three methods use different amounts of output data to construct the index of output growth, as shown in Table 2. The top row (actual) shows the raw count of output delivered each year. A large proportion of data are lost by the requirement that output categories are constant over two successive years, as shown in the second row (Method A). For instance, when comparing growth between t_0-t_1, 20 units of output relating to category J are lost in t_0 and 3000 units relating to category C are lost in t_1. Categorisation changes mean that almost half of the volume of output is lost at times t_1 and t_2. Data loss is less severe under Method B, shown in row 3. This is because mapping of output categories (E=F+G; J=H+I) allows data for these categories to be preserved. Nevertheless output relating to categories C and D, cannot be mapped and are, therefore, omitted from the overall index. Finally, no output data are lost under Method C.

Estimates of output growth derived from applying each method are shown in Table 3. The actual growth rate is shown in the top row; this captures a pure volume effect, as costs are held constant across output categories and over time. The estimates of output growth under both Methods A and B are markedly different to the actual growth rate. This is entirely due to their selective use of data. In contrast, output

[2] Analogously, for the Paasche index which uses cost weights in the year t+1, the problem arises when previous output categories are no longer used.

growth under Method C is identical to actual growth. It is straightforward to apply any other general price deflator (eg as a sensitivity analysis) or, indeed, to use estimates specific to each category, if these are available. Thus the assumptions required for imputing costs are both testable and more benign than those required for mapping and for dealing with activities for which no mapping is possible.

2.1.2 Incorporating quality aspects.

Using costs to weight different outputs implies that they reflect the marginal value that society places on each of these outputs. This holds only under certain assumptions, particularly that health care resources are allocated in line with societal preferences (ie the health system is allocatively efficient). Although this condition is unlikely to be met, at least cost-weights have the advantage that they are reasonably easy to obtain. But in non-market contexts where costs do not reflect demand, they may inaccurately reflect the relative value of outputs. Deviation between costs and values is likely to widen when outputs are subject to changes in quality.

One way to account for quality change is to scale cost-weighted output according to quality change. In general terms, the quality-adjusted cost weighted output index can be written as Dawson et al. (2005):

$$I^{cq} = \frac{\sum_{j=1}^{J} x_{jt+1} c_{jt} \left[\frac{v_j q_{jt+1}}{v_j q_{jt}} \right]}{\sum_{j=1}^{J} x_{jt} c_{jt}} \quad \textbf{(4)}$$

Where q_j represents a unit of quality for output *j*, and v_j is the value of this unit of quality. Incorporating measures of quality in an output index is hampered by a lack of consensus about how to define and measure quality and by the availability of data (Smith and Street 2007). It also has to be recognised that there is no general definition of the quality of health care activities and that the form of quality adjustment is likely to be specific to particular types of output.

In populating this index for the NHS, we allow the characterisation of quality to vary across healthcare settings, partly because activities in different settings have different quality characteristics and partly because the available data differ by setting. Of total health care activities $j=1\ldots J$, we denote activities performed in the secondary care sector as $j=1\ldots S$; in outpatient departments as $j=S_{+1}\ldots O$, in primary care as $j=O_{+1}\ldots P$, and in all other settings as $j=P_{+1}\ldots J$.

The quality adjustment that applies to secondary care activities takes the following form (Dawson, et al. 2005):

$$I^{cq\_S} = \frac{\sum_{j=1}^{S} x_{jt+1} c_{jt} \left(\frac{a_{jt+1} - k_{jt+1}}{a_{jt} - k_{jt}} \right) \frac{\left[\frac{\left(1 - e^{-r_Q LE^{in}_{jt+1}}\right)}{r_Q} - \frac{\left(e^{r_w w_{jt+1}} - 1\right)}{r_w} \right]}{\left[\frac{\left(1 - e^{-r_Q LE^{in}_{jt}}\right)}{r_Q} - \frac{\left(e^{r_w w_{jt}} - 1\right)}{r_w} \right]}}{\sum_{j=1}^{S} x_{jt} c_{jt}} \qquad \textbf{(5)}$$

In essence this quality adjustment attempts to measure the changes from one period to another in quality-adjusted life years (QALYs) and in how long patients wait prior to hospital admission.

Making the QALY adjustment is not straightforward simply because information on the QALYs gained from treatment is unavailable – neither is the change in each patient's health status measured nor is it known for long this change is experienced. To address this information deficit, we use three types of information to create the equivalent to a QALY profile for each type of output. Firstly, we account for whether or not the patient survives treatment by measuring the 30-day post discharge survival rates for each output (a_{jt}). Secondly, we measure the ratio of average health status before and after treatment (k_{jt}). Thirdly, we capture the duration of treatment benefit by estimating the life expectancy associated with each output (LE_j^{in}) by considering the age and gender profiles of patients having each treatment at each time period. r_Q is the discount rate applied to future life years.

The final terms in equation (5) capture changes in waiting times for each output, w_j, in recognition of the welfare loss associated with not being treated immediately. This formulation implies that the marginal disutility of waiting increase as the delay increases. This is similar to charging interest on the cost of waiting, captured by the discount rate r_w. Waiting time is measured as the 80th percentile of the waiting time distribution for each type of treatment. The choice is driven by the realisation that reductions in relatively long waiting times confer benefits on all patients by reducing the risk of having to face a very long wait.

Patients also experience increasing disutility the longer they have to wait for an outpatient appointment. Changes in outpatient waiting times are captured as a quality adjustment to outpatient output as follows:

$$I^{cq\_out} = \frac{\sum_{j=S_{+1}}^{O} x_{jt+1} c_{jt} \left[LE^{out} - \frac{\left(e^{r_w w_{jt+1}} - 1\right)}{r_w} \Big/ LE^{out} - \frac{\left(e^{r_w w_{jt}} - 1\right)}{r_w} \right]}{\sum_{j=S_{+1}}^{O} x_{jt} c_{jt}} \qquad \textbf{(6)}$$

This is a simplified form of the quality adjustment for waiting times for hospital admission. The mean waiting time is used as a value for w_j and LE^{out} captures life expectancy, which is taken as a constant of 26 years.

NHS output performed in the primary care setting is also adjusted to capture improvements in the control of cholesterol and high blood pressure (Derbyshire, et al. 2007). The quality adjustment applying to primary care activities takes the form:

$$I^{cq\_P} = \frac{\sum_{j=O_{+1}}^{P} x_{jt+1} c_{jt} \left[\frac{v_j B_{jt+1}}{v_j B_{jt}} \right]}{\sum_{j=O_{+1}}^{P} x_{jt} c_{jt}} \qquad \textbf{(7)}$$

Where B measures the proportion of consultations for patients with coronary heart disease, stroke and hypertension in which blood pressure is recorded as being below the threshold levels specified in the Quality and Outcomes Framework (Department of Health 2004). v_j reflects the value of consultation where low blood pressure is recorded relative to a consultation where the reading is above the threshold.

For NHS output administered in all other settings (eg community services, ambulance, NHS Direct), no quality adjustors are currently available. Hence, for these activities we resort to using a simple cost weighted output index as shown in equation (3).

2.2 Indices of input growth

Inputs into the health care system consist of labour, intermediate goods and services and capital elements. Growth in the use of these so called "factors of production" can be calculated directly or indirectly. A direct measure of input growth can be calculated when data on the volume and price of inputs are available such that input growth is measured as:

$$Z^D = \frac{\sum_{n=1}^{N} z_{nt+1} \omega_{nt}}{\sum_{n=1}^{N} z_{nt} \omega_{nt}} \qquad \textbf{(8)}$$

Where z_{nt} is the volume of input of type *n* at time *t* and ω_{nt} is the price of input type *n* at time *t*. In practice it is difficult to calculate this index because volume data are rarely available. The alternative is to employ an 'indirect' measurement approach that relies on expenditure data. Expenditure is driven by both the volume and price of inputs:

$$E_1 = z_1 \times \omega_1 \qquad \textbf{(9)}$$

Where E_1 is the total expenditure on input 1, z_1 is the volume of input 1 and ω_1 is its price. The growth in total expenditure from one period to the next is driven by changes in both the volume and price of inputs:

$$\Delta E_1 = \frac{E_{1t+1}}{E_{1t}} = \frac{z_{1t+1}\omega_{1t+1}}{z_{1t}\omega_{1t}} \quad \textbf{(10)}$$

Thus, expenditure growth comprises both a price effect and a volume effect. To isolate the volume effect it is necessary, therefore, to convert 'nominal' monetary values into 'constant' or 'real' expenditure using a deflator π. This deflator reflects the underlying trend in prices for the inputs in question, such that $\omega_{nt} = \pi_n \omega_{nt+1}$. If these are available, the input growth index can be specified as:

$$Z^{Ind} = \frac{\pi_n E_{nt+1}}{E_{nt}} = \frac{\sum_{n=1}^{N} z_{nt+1}\pi_n \omega_{nt+1}}{\sum_{n=1}^{N} z_{nt}\omega_{nt}} = \frac{\sum_{n=1}^{N} z_{nt+1}\omega_{nt}}{\sum_{n=1}^{N} z_{nt}\omega_{nt}} = Z^D \quad \textbf{(11)}$$

Provided that accurate price deflators are used, and that accurate data are available about the volume and price of each factor, direct and indirect measurements of input growth are equivalent, as shown in equation (11). The indirect method is, however, likely to be a more comprehensive measure of resource use, if price deflators are sufficiently disaggregated to take into account changes in the mix of inputs and reflect full and actual costs (Atkinson 2005).

The measurement of capital input growth poses special challenges. Firstly, the contribution of capital in any given period will be a function partly of current expenditure on the asset (E_{nt}) and will also draw on previous investments $E_{nt-1}, E_{nt-2},...$ The question is then; what proportion of past expenditure is devoted to current production? And secondly, what proportion of current expenditure represents investment that will be drawn upon in the future?

The proportions reflect the rate of decay[3] of the asset, denoted θ_n. The decay rate that applies to each period's expenditure relates to how old the asset is (its vintage) at the time that (some proportion of) it is used (Oulton, 2001 and 2005; Oulton and Srinivasan 2003). Hence, if we consider an asset deployed across just two periods our indirect input growth index would take the following form:

$$Z^{Ind} = \frac{\pi_n \left[\theta_{nt}E_{nt} + \theta_{nt+1}E_{nt-1}\right]}{\theta_{nt}E_{nt-1} + \theta_{nt+1}E_{nt-2}} \quad \textbf{(12)}$$

Where $\theta_{nt}E_{nt}$ captures the proportionate contribution of current expenditure and $\theta_{nt+1}E_{nt-1}$ captures the proportionate contribution of expenditure in the previous period to current production. In the denominator, $\theta_{nt}E_{nt-1}$ captures the proportionate contribution of the previous period's expenditure on production in that period and $\theta_{nt+1}E_{nt-2}$ captures the proportionate contribution of expenditure two periods earlier to production in the previous period. Equation (12) needs to be extended back over all the periods from which past investments are drawn in the current use of the asset.

[3] The rate of decay measures how the contribution to productive output from the asset declines over time.

2.3 A note on prescribing

It has been conventional in the UK national accounts to include medicines prescribed in primary care both as an input into the production of health care and as health care outputs in their own right. This contrasts to how medicines delivered in the hospital sector are dealt with, which are considered solely as an input. The justification for including drugs prescribed in the primary care setting is that, first, GPs add value through prescribing - otherwise all licensed drugs would be available over the counter - and, second, the value GPs add to health care output is not reflected in the assumption that the wage rate approximates the marginal product of GPs (Dawson, et al. 2005).

There would be no need to appeal to this justification if the unit of health care output was measured as the number of patients treated in primary care weighted by their health gain (Dawson, et al. 2005). Pharmaceuticals would then be counted consistently as an input. If prescribing more expensive drugs turned out to be cost effective in improving health outcomes, this would appear as a productivity increase provided that changes in health status are measured accurately and routinely. However, this preferred form of the output index cannot be calculated because measures of health outcomes are not available for the majority of treatments. Current accounting practice, therefore, is a compromise in the absence of the requisite data.

3. Data sources and description

3.1 Data sources - outputs

Table 4 provides a summary of data sources and data summary statistics for NHS outputs. The hospital episode statistics (HES) are the prime source for identifying output growth in the provision of inpatient and day case services to NHS patients. HES covers all medical and surgical specialities and includes private patients treated in NHS hospitals. In addition, HES captures admitted patient care funded by the NHS but provided by the independent sector – although the quality of data from some independent sector providers is poor (Healthcare Commission 2007; Mason, et al. 2009).

HES now comprises over 15 million patient records each year. Records are stored according to the financial year in which the period of care finished and each includes a number of data fields, containing demographic data (e.g. age, gender), clinical information (e.g. diagnoses, procedures performed) and details of the hospital and specialty where the patient received treatment. We are also able to link HES data to death registry records, so deaths following discharge can be measured.

A HES record is defined as a ‘finished consultant episode’, which is the time that a patient spends under the care of a single consultant. During their course of treatment a patient may be treated by more than one consultant and may be transferred to another hospital, with a new record being created each time this happens. Therefore, some data manipulation is required in order to measure the total care received by a patient from hospital admission to the end of their treatment in the secondary care sector. We improve on the method described by Lakhani et al. (Lakhani, et al. 2005) to identify continuous inpatient spells (CIPS) which track patients when transferred between consultants and hospitals as part of their care pathway. Specifically, we allow for the linking of records across successive financial years. A full description of the algorithm used to identify CIPS and its new characteristics can be found in Castelli et al. (2008).

The Reference Costs capture data about activities conducted in many health care settings. First introduced in 1997 coverage has increased to include progressively more NHS output conducted in NHS Trusts, Primary Care Trusts (PCTs), Personal Medical Services[4] (PMS) and more recently non-NHS providers. Data are reported using various means of describing a unit of output, including HRGs, contacts, bed days, and number of tests.

Data collection has undergone substantial revisions over time, from the introduction of previously uncounted output to changes in the structure of the classifications system used (eg revisions to outpatient and community care groups), to re-categorisations (disaggregation/re-aggregation) of previously counted output. These changes make it challenging to construct a consistent output series; thus, requiring some means of dealing with re-categorisation, as discussed in Section 2.1.1.

Estimates of Primary care consultations are based on data collected as part of the QRESEARCH[5] project. This data source replaces previous estimates of primary care output derived from the General Household Survey, which is known to be deficient in measuring changes in consultation rates from year to year (Atkinson 2004). QRESEARCH is a large consolidated database derived from the anonymised health records of over 9 million patients. The data currently come from 554 general practices in the UK using a common clinical computer system. These data are used to derive estimates of primary care output for England as a whole (Fenty, et al. 2006). [6]

The volume and cost of prescribing is derived from Prescription Pricing Authority (PPA) data. The PPA data are collected in order to remunerate pharmacists and dispensing GPs. It is therefore a reliable and comprehensive measure of the volume of prescriptions dispensed, which can be disaggregated to product type (item). Finally, information on the volume of ophthalmic and dental services is derived from the Information Centre (The Information Centre 2007a; The Information Centre 2007b).

3.2 Data sources – inputs

Direct measurement of input growth is feasible for only two input types, volume data being available from the PPA about prescriptions dispensed and from the workforce census about the number of NHS staff employed. But the workforce census provides a partial indication of labour input, as overtime working by NHS staff and the contributions of agency staff are not captured. In view of this, we use the financial returns made by all NHS organisations, and - for all inputs but prescriptions - apply the indirect method to calculate input growth by deflating expenditure on each input type, as summarised in table 1.5.[7]

The financial returns contain two forms of information about capital expenditure: current outlays on equipment and past expenditure reported as depreciation on assets. We make assumptions about what proportion of current expenditure is employed in the current period (Street and Ward 2009). It is likely that estimates of depreciation are inaccurate, particularly as infrequent outlays are likely to be under-represented. But they have the advantage that they are based on the particular composition and vintage of capital in each organisation. We believe that the use of these data is preferable to current national accounting practice, whereby capital growth is based on an historical data series that is not specific to the

[4] "PMS is a locally-agreed alternative to General Medical Service (GMS) for providers of general practice. Legislation has allowed for PMS since 1997 but it is only in recent years that the number of practices choosing PMS has grown rapidly. Now almost half of general practices have PMS agreements". http://www.bma.org.uk/ap.nsf/content/pmsagreements0904 (last accessed 1st July 2008).

[5] http://www.qresearch.org/Public/WhatIs.aspx

[6] http://www.ic.nhs.uk/webfiles/publications/Consultations%20Report%2015%20financial%20year%20April2008.xls (last accessed 19/06/09).

[7] The use of direct measurement approach to estimate labour input is explored in Street and Ward (2009).

NHS and which the ONS recognises as being imperfect (Lee 2006). Price deflators appropriate to each type of capital asset are applied, in recognition that changes in price and quality vary according to the asset in question (Wallis 2005).

3.3 Data trends

Table 4 summarises the main elements used in the construction of the output index, full details of which are in Castelli et al. (2008). Elective, emergency and mental health activity and survival rates have all improved over time and waiting times have fallen, all of which contribute positively to growth. Over the same period the NHS has been extending treatment to older patients. If the age profile of NHS patients increases more rapidly than the improvements in population life expectancy this will lead to a dampening of output growth. Output growth depends, then, on the net effect of these various conflicting influences.

Outpatient activity has increased steadily over the time period considered. However, some caveats need to be drawn as the Reference Costs data collection has undergone substantial changes with an increase in the number of categories over time, and a major overhaul of the categorisation architecture for the 2006/07 collection exercise. Outpatient waiting times have experienced a continuous reduction over the four years, falling from 8 weeks in 2003/04 to 6 weeks in 2006/07.

Community care services comprise all NHS activity provided to NHS patients in a community setting, such as nursing services, midwifery visits, and therapy services. Data collection for these services has undergone major changes to the classification system, which we account for by imputing costs when these are unavailable. Other NHS activity encompasses all remaining health care output such ambulance services, A&E activity, renal dialysis, NHS Direct, Ophthalmic services. There has been a major expansion in data collection for community and 'other' NHS activity, with PCTs making increasingly detailed returns since 2003/4. This drives much of the 'growth' in these activities, and in the next section we construct output indices with and without these activities.

Primary care activity (eg. GP and nurse consultations) has recorded a modest increase in the first two years of the time period considered, followed by a rapid increase in 2005/06 before levelling off in the last year. Finally, there have been year-on-year increases in the volume of prescriptions dispensed.

Table 5 reports expenditure in constant terms for each of the main factors of production. Full details are available in Street and Ward (2009). The level of input growth in the NHS as a whole has averaged 4.7 % per year. Growth in labour input has varied over this time: after an initial increase in the number of staff recruited into the NHS numbers have recently began to level off. This is mainly due to reduced reliance on agency staff. The volume of intermediate inputs and capital has also increased year on year.

4. Results

Tables 6 and 7 show respectively the unadjusted and the quality-adjusted cost weighted output growth figures by sector and for the NHS as a whole.

- The cost weighted output index, both unadjusted and quality-adjusted, for the hospital sector (inpatient and day case) is positive throughout the period, averaging 3.62% and 6% per year respectively. Improvements in survival rates and waiting times account for 2.4% of the difference in the annual growth figures.
- Growth in outpatient activity averages 4.39% in the unadjusted index and 4.5% per annum in the quality-adjusted one. Outpatient waiting times have improved in the time period under investigating

leading to a positive effect on the growth rate. Further, there is a fall in output growth between 2005/6 and 2006/7: even though more patients were seen, they were of lower ‘complexity’ than previously.

- Growth in mental health care has averaged between 8.6% and 8.7%, according to whether or not activity has been quality-adjusted. The majority of the output growth is recorded between 2003/04 and 2004/05 and is largely due to a shift towards more costly types of activity. The quality adjustment is only applied to the mental health inpatient data, which represent about 1% of outpatient mental health services. This explains their limited contribution to the overall figure.
- Growth in primary care consultations and prescriptions has averaged between 5.26% and 5.45% over the full period, respectively for the unadjusted and the quality-adjusted output growth indices. Quality adjustment is measured in terms of improvements in the management of blood pressure by GPs contributing for about 0.5% of the average annual growth rate of the primary care consultations only.
- The very high growth rate for community services between 2003/4 – 2004/5 is driven by much improved data collection intended to ensure that these activities were recorded comprehensively. Hence the figure of 315% is more a reflection of data collection than it is of true output growth.
- Data collection has also improved over time with respect to activities categorised under “all other NHS activity”. Hence, the growth rates will reflect both changes in data collection and actual increases in output.
- The growth rates for the Total NHS are aggregated from the figures across all settings, and suggest annual output growth of 14% per year. As mentioned above, though, this figure is contaminated by improvements in data collection.
- The final row presents growth rates for the secondary and primary care sectors only. This captures activity in hospitals, outpatient departments, in mental health and in primary care. Data collection has been consistent and comprehensive over the full period. This means that they measure output growth accurately. These imply that output growth has averaged 5.5% per annum.

Table 8 details the proportionate split of labour, capital and intermediate inputs by organisational type. The estimates shown are based on indirect measurement of each factor of production. The secondary care sector (comprising hospitals, FTs and ambulance trusts) is highly labour intensive, accounting for on average 72 percent of total inputs. This proportion has changed over time, particularly as capital investment has increased. There is an even mix between expenditure on labour and intermediate inputs by PCTs, mainly because purchases from non-NHS bodies are included in the latter category. Capital accounts for around 6% of NHS inputs, a higher proportion than the 2% estimate used by the ONS (Lee 2006). This may still under-represent capital inputs, if these are not fully documented in organisational financial returns.

5. NHS productivity

Table 9 reports indices of output and input growth, taking 2003/04 as the base year, followed by year-on-year estimates of productivity growth. Two sets of estimates are presented, one for the secondary and primary care sectors only, and the other for the entire NHS. The latter estimates are contaminated by improved data collection relating particularly to community services during 2003/04-2004/05.

For secondary and primary care only, outputs have grown faster than inputs, implying positive productivity growth. For the NHS as a whole productivity grows very strongly initially. This is driven largely by improved data collection in the community care sector but levels off in subsequent years. In the latter years, outputs show a faster growth rate compared to inputs resulting in positive productivity growth.

We supplement these figures with estimates of output, input and productivity growth taken from our earlier report (Dawson, et al. 2005) to construct consistent series dating from 1998/99 for the primary and secondary sectors. These series are presented in Figure 1.

Since 1998/9 there has been strong input growth, particularly after 2000/1. Staff received new pay awards. Recruitment increased, in part to satisfy the European Working Time Directive. There was greater investment in equipment and buildings. Meanwhile output growth lagged behind input growth. This is unsurprising. The directive placed limits on working hours, reducing the number of patients per doctor, and facilities invested in take time to build. The year-on-year increases in patients treated meant that output growth averaged more than 3.7% per year up to 2003/4. But the net effect was negative productivity growth between 1998/9 and 2003/4.

This has since changed. Output in the primary and secondary care sectors has continued to rise, but at a faster rate, averaging 5.5% per year between 2004/5 and 2006/7. Not only are more patients being treated, but the quality of the care they receive has been improving. Waiting times have been falling, both for outpatient appointments and for admission to hospital, survival rates have been improving for patients admitted to hospital whether as electives or non-electives, and improved disease management in primary care has led to reductions in blood pressure for patients suffering chronic heart disease, stroke and hypertension. There has also been a slowdown in input growth since 2003/4. This is primarily the result of a levelling off in staff recruitment and less reliance on the use of agency staff. As a consequence recent productivity growth has been positive or, at least, constant. This means output growth is at least as high as the growth in inputs.

5. Conclusions

Our indices of NHS output and input growth are designed to be comprehensive, capturing as far as possible all the activities undertaken for NHS patients by both NHS and non-NHS providers. This contrasts with most indices that are based on a 'basket' of activities that are deemed to be representative of the whole. We analyse information about every patient treated in hospitals and outpatient departments and about every prescription dispensed in primary care and improve on methods used to define the patient pathway in secondary care, tracking patients across providers and financial years. We also construct a comprehensive index of input growth, making use of financial returns from every NHS organisation and enabling us to improve on estimates of capital utilisation.

We address a major practical challenge that arises in the NHS because of periodic wholesale revisions to the classification systems used to describe output categories. Reference Cost categories used to describe outputs in community care were completed re-defined after 2005/6. Although not included in this paper, in 2007/8 there was a complete revision of the way that hospital output is defined, with the move from version 3.5 to version 4 HRGs. If we had relied on traditional methods, output growth between 2005/6 and 2006/7 would be based solely on hospital and primary care activity and, looking ahead, output growth between 2006/7 and 2007/8 would be based solely on non-hospital activity. Clearly, comparisons of output growth over the full period would be rendered virtually meaningless, with only primary care activity being included throughout.

Significant improvements to data collection in some health care settings make it difficult to isolate true output growth from improved reporting. Conservatively, therefore, we focus on those settings where data collection has been consistent as better representing actual growth for the period we consider. We conclude that, recently, growth in inputs is matched by growth in outputs, suggesting that the NHS is operating under constant returns: we get out what we put in.

What might our analysis mean as we enter a more resource-constrained period? There is a danger that across-the-board 'efficiency' savings may translate simply into commensurate reductions in the number of patients being treated or in the quality of the care they receive. To guard against this, it will be important to examine variations in productivity in different parts of the country so that efforts can be targeted to where most gains are to be made. In future work we shall also explore the possibility of producing sub-national (Strategic Health Authority) estimates of input, output and productivity growth. These will not only be interesting in their own right but will also aid our understanding of the drivers of productivity.

Acknowledgements

Particular thanks are due to Mark Chandler, Keith Derbyshire, Warren Goldthorpe, Hugh Gravelle, Katherine Mills and Panos Zerdevas. The project was funded by the Department of Health in England as part of a programme of policy research at the Centre for Health Economics, University of York. The views expressed are those of the authors and may not reflect those of the funder.

References

Atkinson, T. (2004) *Atkinson Review: Interim Report*, London: The Stationary Office.

— (2005) *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*, Basingstoke: Palgrave Macmillan.

Castelli, A., Laudicella, M. and Street, A. (2008) 'Measuring NHS output growth', York: Centre for Health Economic research paper 43.

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*, Chicago: University of Chicago Press.

Dawson, D., Gravelle, H., O'Mahony, M., Street, A., Weale, M., Castelli, A., Jacobs, R., Kind, P., Loveridge, P., Martin, S., Stevens, P. and Stokes, L. (2005) *Developing new approaches to measuring NHS outputs and productivity, Final Report*, York: Centre for Health Economic research paper 6.

Department of Health (2004) *Quality and outcomes framework: guidance – updated August 2004*, London: Department of Health.

Derbyshire, K., Zerdevas, P., Unsworth, R. and Haslam, M. (2007) *Further developments in measuring quality adjusted healthcare output*, Leeds: Department of Health.

Eurostat (2001) *Handbook on price and volume measures in national accounts*, 2001 Edition: Luxembourg: Office for Official Publications of the European Communities.

Eurostat/Commission of the European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and Bank, W. (1993) *System of National Accounts 1993*: European Communities/EUROSTAT, IMF, OECD, United Nations and World Bank.

Fenty, J., Coupland, C., Hippisley-Cox, J. and Gravelle, H. (2006) *Determinants of consultation rates over time: Implications for estimating the national volume of consultations.* , Nottingham: QRESEARCH University of Nottingham.

Healthcare Commission (2007) *Independent sector treatment centres: a review of the quality of care*, London: Commission for Healthcare Audit and Inspection.

Lakhani, A., Coles, J., Eayres, D., Spence, C. and Rachet, B. (2005) 'Creative use of existing clinical and health outcomes data to assess NHS performance in England: Part 1-performance indicators closely linked to clinical care', *British Medical Journal* 330: 1426-1431.

Lee, P. (2006) *Public Service Productivity: Health*, London: Office of National Statistics.

Mason, A., Miraldo, M., Siciliani, L., Sivery, P. and Street, A. (2009) *Establishing a fair playing field for Payment by Results*, York: Centre for Health Economic research paper 39.

Office for National Statistics (2009) *Health Care Output 1995 - 2007*: Office for National Statistics.

Oulton, N. (2001) 'Measuring capital services in the United Kingdom', *Bank of England Quarterly Bulletin* Autumn: 295-309.

— (2005) *Ex post versus ex ante measures of the user cost of capital*, London: Centre for Economic Performance.

Oulton, N. and Srinivasan, S. (2003) *Capital stocks, capital services, and depreciation: an integrated framework*, London: Bank of England.

Pike R. (2008) 'Measuring Inflation', *Economic & Labour Market Review* 2(7): 59-60.

Smith, P. C. and Street, A. (2007) *Workshop on the measurement of non-market output in education and health*, York: Centre for Health Economics research paper 23.

Street, A. and Ward, P. (2009) *NHS Input and Productivity Growth 2003/4 - 2007/8. Research Paper 47*, York: University of York. Centre for Health Economics.

The Information Centre (2007a) *General Ophthalmic Services: Activity statistics for England and Wales*, Leeds: The Information Centre.

— (2007b) *NHS Dental Activity and Workforce Report England: 31 March 2006*, Leeds: The Information Centre.

Wallis, G. (2005) 'Estimates of the volume of capital services', *Economic Trends* 624: 42-51.

Table 1 –Illustrative sample of categories

Categories	t_0		t_1		t_2		t_3	
	activity	*cost*	*activity*	*cost*	*activity*	*cost*	*activity*	*cost*
A	750	1	820	1	700	1	650	1
B	1000	1	1500	1	2200	1	3500	1
C	0	0	3000	1	4200	1	0	0
D	20	1	0	0	0	0	4500	1
E (=F+G)	1500	1	1700	1	0	0	0	0
F	0	0	0	0	900	1	1800	1
G	0	0	0	0	1200	1	2400	1
H	250	1	620	1	0	0	0	0
I	3000	1	3250	1	0	0	0	0
J(=H+I)	0	0	0	0	4000	1	2000	1
Total activity	6520		10890		13200		14850	
Actual growth	-		67.02%		21.21%		12.50%	

Table 2 Data used in the output index under each method

Total Activity count	t_0 - t_1		t_1 - t_2		t_2 - t_3	
	t_0	t_1	t_1	t_2	t_2	t_3
Actual	6,520	10,890	10,890	13,200	13,200	14,850
Method A	6,500	7,890	5,320	7,100	9,000	10,350
Method B	6,500	7,890	10,890	13,200	9,000	10,350
Method C	6,520	10,890	10,890	13,200	13,200	14,850

Table 3 Output growth, Laspeyres index

Index	t_0 - t_1	t_1 - t_2	t_2 - t_3
Actual growth	67.02%	21.21%	12.50%
Method A growth	21.38%	33.46%	15.00%
Method B growth	21.38%	21.21%	15.00%
Method C growth	67.02%	21.21%	12.50%
Difference	0.00%	0.00%	0.00%

Table 4 NHS activity sources and data description

Data Source	NHS activity	Year 2003/04	2004/05	2005/06	2006/07
Hospital Episode Statistics (HES)	**Hospital output**				
	Elective and day cases				
	Volume of activity	6,401,519	6,433,933	6,864,612	7,194,697
	Activity weighted average unit cost	937	1031	1041	1036
	Mean 30-day post discharge survival rate	0.9936	0.9938	0.9947	0.9951
	Mean age	53.2	53.6	53.9	54.4
	Mean life expectancy	23.54	23.75	23.65	23.59
	80th percentile waiting times	118.7631	103.9977	94.6607	88.8257
	Emergencies				
	Volume of activity	5,723,817	6,009,802	6,291,117	6,363,388
	Activity weighted average unit cost	1126	1210	1241	1244
	Mean 30-day post discharge survival rate	0.9492	0.9516	0.9549	0.9565
	Mean age	41.4	41.6	41.6	41.6
	Mean life expectancy	33.8244	34.1070	34.2516	34.5762
	Mental Health inpatient				
	Elective and day cases				
	Volume of activity	47,384	45,624	41,439	38,408
	Activity weighted average unit cost	676.32	688.58	673.01	655.86
	Mean 30-day post discharge survival rate	0.9778	0.9772	0.9801	0.9815
	Mean life expectancy	29.71	30.09	30.01	30.60
	80th percentile waiting times	159.88	40.35	264.80	256.83
	Emergencies				
	Volume of activity	120,789	123,983	120,203	115,560
	Activity weighted average unit cost	936.88	1,012.07	1,012.17	1,012.17
	Mean 30-day post discharge survival rate	0.9674	0.9696	0.9722	0.9738
	Mean life expectancy	28.06	28.74	28.89	29.03
Reference Costs	**Mental Health**				
	Volume of activity	15,168,410	16,389,891	17,738,894	19,259,205
	Activity weighted average unit cost	151.39	164.36	169.55	166.82
	Outpatient				
	Volume of activity	50,205,073	52,724,302	60,541,477	63,453,507
	Activity weighted average unit cost	98.10	105.63	102.63	92.67
	Mean waiting time (weeks)*	8.32	7.41	6.53	5.94
	Community care				
	Volume of activity	31,342,436	75,673,792	85,092,838	83,895,139
	Activity weighted average unit cost	88.64	39.00	37.88	39.52
	Other NHS activity				
	Volume of activity	196,620,122	230,116,689	271,833,567	338,198,807
	Average unit cost	Acitivity specific			
NHS IC	**WiC, NHS Direct, ophthalmic and dental serv.**				
	Volume of activity	56,753,664	60,733,378	46,111,030	45,536,081
	Average unit cost	Acitivity specific			
QResearch	**Primary Care**				
	Volume of activity (000 contacts)	262,100	265,600	283,100	293,000
	Activity weighted average unit cost**	18.29	20.46	20.65	24.53
PPA	**Prescriptions**				
	Volume of activity (000 prescriptions)	649,155	681,446	722,412	752,503
	Activity weighte average unit costs	10.44	10.52	9.87	9.78

* Data source: DH – performance website; ** Data source: PSSRU

Table 5 NHS inputs sources and data description, (£000s)

Data Source	NHS inputs	Year			
		2003/04	2004/05	2005/06	2006/07
Financial Returns *(NHS Hospital Trusts, PCTs, Consolidated FT accounts)*	**Labour**				
	NHS Staff	57,223,616	59,791,671	61,934,262	61,910,576
	Agency Staff	3,077,825	2,785,714	2,456,212	1,894,629
	Deflator*	NHS Pay Index			
	Intermediary Inputs				
	Intermediary Inputs	22,628,646	23,521,254	26,945,141	29,177,344
	Deflator*	Pay and Prices deflator, FHS deflator			
	Capital Inputs				
	Capital Inputs	5,817,249	6,004,404	6,505,958	7,580,696
	Deflator**	Specific capital assets deflators (MM17 Price Index)			
	Other Inputs				
	Other inputs	7,906,228	9,430,647	10,474,193	10,111,700
	Deflator*	NHS Pay Index, Pay and Prices deflator			

* Data source: DH; ** Data Source: ONS.

Table 6 Unadjusted cost weighted output index, Laspeyres index

Sector	2003/04 - 2004/05	2004/05 - 2005/06	2005/06 - 2006/07	Average
Hospital activity	2.56%	5.48%	2.80%	3.62%
Outpatient activity	10.14%	9.87%	-6.86%	4.39%
Mental Health services	11.44%	9.50%	4.83%	8.59%
Primary care consultations & prescribing	4.33%	6.99%	4.47%	5.26%
Community services	315.53%	10.25%	-0.65%	108.38%
All other NHS activity	17.13%	3.14%	22.07%	14.11%
Total NHS	27.88%	6.48%	5.84%	13.40%
Secondary and Primary Care	5.09%	7.06%	1.94%	4.70%

Table 7 Quality-adjusted cost weighted output index, Laspeyres index

Sector	2003/04 - 2004/05	2004/05 - 2005/06	2005/06 - 2006/07	Average
Hospital activity	5.66%	7.48%	4.88%	6.01%
Outpatient activity	10.23%	9.96%	-6.81%	4.46%
Mental Health services	11.83%	9.42%	4.82%	8.69%
Primary care consultations & prescribing	4.51%	7.16%	4.67%	5.45%
Community services	315.53%	10.25%	-0.65%	108.38%
All other NHS activity	17.13%	3.14%	22.07%	14.11%
Total NHS	28.82%	7.11%	6.08%	14.00%
Secondary and Primary Care	6.44%	7.90%	2.30%	5.55%

Table 8 Proportionate split in factors of production

Sector	Year				Average
	2003/04	2004/05	2005/06	2006/07	
Trusts					
Labour	73%	73%	72%	70%	72%
Intermediate	20%	19%	21%	22%	20%
Capital	7%	7%	7%	9%	8%
PCTs					
Labour	50%	51%	48%	45%	49%
Intermediate	46%	45%	47%	50%	47%
Capital	4%	4%	4%	5%	4%
NHS					
Labour	58%	57%	54%	53%	55%
Intermediate	37%	38%	40%	41%	39%
Capital	6%	5%	5%	6%	6%

Table 9 Output and input growth indices (2003/04=1) and Productivity growth

	Year		
	2004/05	2005/06	2006/07
Output growth			
Secondary and primary care	1.06	1.49	1.78
Total NHS	1.29	1.38	1.46
Input growth			
Secondary and primary care	1.07	1.13	1.15
Total NHS	1.06	1.14	1.16

	Year		
	2003/04 - 2004/05	2004/05 - 2005/06	2005/06 - 2006/07
Productivity growth			
Secondary and primary care	-0.06%	1.64%	0.40%
Total NHS	21.14%	-0.08%	3.85%

Figure 1: Growth in outputs, inputs and productivity, 1998/9 – 2006/7