PRELIMINARY DRAFT FOR HESG GLASGOW JULY 2004
Check with authors before citing

Developing new approaches to measuring NHS outputs and productivity

Diane Dawson[1], Hugh Gravelle[**], Paul Kind[*], Mary O'Mahony[***],
Andrew Street[*], Martin Weale[***]

Abstract

HM Treasury and the Department of Health have a short term (three year) Public Service Agreement that requires a 2% per annum increase in productivity of the National Health Service (NHS). It is intended that this will be achieved through cost reductions and quality enhancements. The aims of the paper are to identify the conceptual and practical challenges in constructing measures of NHS inputs and outputs so that expenditure change can be disaggregated into price increases, treatment expansion, treatment substitution and productivity. Such measures require data on inputs, outputs, outcomes and a suitable set of prices (weights).

Methods. We describe the data requirements for measuring NHS productivity. We examine the arguments for measuring outcomes as well as output, discuss when unit costs are a valid output weight and the implications of technological change and service innovations, changing patient mix, absent prices, imperfect input markets, and imperfect agency. We present some simple calculations based on a small number of HRGs for which we have data on outputs, outcomes (QALY gains, waiting times, mortality) and unit costs to illustrate the impact of different index number formulations on the rate of growth of NHS output.

Possible issues for HESG discussion. QALYs to measure NHS outcomes? What other outcomes are relevant (waiting times, process utility)? Weights for outputs/outcomes: unit costs, contingent valuations, private sector prices? What is the output of primary care? What is the relationship between CEA and output measurement? Should we allow for unpaid carer and patient time and if so how should it be valued? Suggestions for data?

[1]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

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 productivity in the UK health sector using a methodology commonly applied to other sectors has found that productivity growth in health compares favourably with the rest of the economy (O'Mahony and DeBoer, 2002). This discrepancy shows that the method of measurement is central. Such concerns led to the setting up of the Atkinson Commission to review the measurement of public sector productivity.

UK studies both of expenditure per unit of activity and of productivity have found it problematic to incorporate measures of quality change. For some sectors of the economy this is likely to have been an important source of productivity change and there is an expectation that it may be true in health. 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 may 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 DH has commissioned a team from CHE and NIESR to investigate methods of measuring outputs and productivity in the NHS and to examine the feasibility of developing a workable measure of productivity change for the NHS. The 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 1 subject to data availability reviewed in Stage 2.

In this paper we set out some of the methodological and data problems which arise in attempting to measure NHS outputs and productivity and present a few examples, based on a small number of outputs, to illustrate how existing data sources may be combined to produce new estimates of output growth. Section 2 discusses why one would want to measure productivity, distinguishes between activities, outputs and outcomes, and sets out the data requirements for constructing the type of index numbers typically used to measure productivity growth. Section 3 discusses the problems with constructing an index of NHS output and Section 4 presents an illustration of the ways in which different types of data on outcomes, outputs and costs can be combined in an output index using data on 9 HRGs. The focus of the paper is on output indices. Space constraints ruled out consideration of the problems in constructing input indices for the NHS. A brief account of these is in Dawson et al (2004).

2 Productivity measurement

2.1 Why measure productivity growth?

Total factor productivity (TFP) is calculated as the difference between the growth rates of measures of output and input. TFP can be interpreted as a measure of technological change. The growth accounting method allocates sources of output growth to changes in factor inputs and TFP. The approach has been used extensively not only to compare performance between sectors of the economy but also to investigate the relative contribution of different inputs to productivity change (Jorgenson et al., 1987, O'Mahony, 1999, Triplett and Bosworth, 2002). In order for TFP to be interpreted as a measure of technological progress it must be assumed that the sector being examined is technically efficient ie it would not be possible to produce more of at least one output from the inputs used. This seems unlikely to be satisfied in the NHS.

Consider Figure 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, B has lower technical efficiency since it is further from its period production frontier. Technical progress has shifted the frontier upward 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 have diminishing returns to scale so that increases in inputs along the frontier reduce productivity. Figure 2 plots NHS productivity growth, as measured by the old CWAI against the growth in inputs. The negative relationship (the larger the rate of input growth the lower the rate of productivity increase) is compatible with diminishing returns (or adjustment costs, or a purely fortuitous association of input growth and reduced productivity, or…).

Productivity growth can be a useful summary statistic but it should not be viewed as a welfare or efficiency measure and it needs supplementing with other data on the NHS. A further justification for attempting to measure productivity is that it will stimulate improvements in NHS information collection and processing which will improve

decision making within the NHS.

2.2 Activities, outputs, outcomes

It is useful to distinguish activities (operative procedures, diagnostic tests,…), 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 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, the costs of accessing the service, the range of choice and quality of hospital food, the surliness of the practice receptionist, the involvement of the patient in decisions about their treatment,… In the measurement of private sector output the focus is on counting outputs rather than the characteristics because of the assumption that the price of the output measures the consumer's marginal valuation of the bundle of characteristics or outcomes from consuming the output. In the absence of explicit monetary prices we may have to pay greater attention to measuring and valuing the outcomes. Similarly in measuring private sector productivity we do not need to concern ourselves with counting activities because these are embodied in the outputs which are produced and sold. In the NHS the lack of final market where consumers buy outputs from producers means that there is frequently only information on the activities (and not always on them).

Suppose for the moment that we can measure NHS outputs. We can define the marginal social value of the output of the NHS over year *t* as:

$$Y_t = \sum_j \sum_k \eta_{kt}\zeta_{kjt}x_{jt} = \sum_j p_{jt}x_{jt} \qquad (1)$$

where x_{jt} is the volume of output j in year t, ζ_{jkt} is the amount of outcome (characteristic) k produced per unit of outcome j in year t, and η_{kt} is the marginal social value of outcome k in year t. Note that we measure Y_t either by placing a value on outcomes or by placing a value on the output which reflects the outcomes it yields:

$$p_{jt} = \sum_k \eta_{kt}\zeta_{jkt} \qquad (2)$$

The choice between these approaches will be determined by the availability of data on p, ζ and η.

The rate of growth of Y_t can be decomposed as a weighted sum of the growth rates of the volume of outputs, the outcomes per unit output and the value of the outcomes:

$$\frac{d\log Y_t}{dt} = \frac{1}{Y_t}\sum_j \sum_k \left(\eta_{kt}\zeta_{kjt}\dot{x}_{jt} + \eta_{kt}\dot{\zeta}_{kjt}x_{jt} + \dot{\eta}_{kt}\zeta_{kjt}x_{jt}\right) = \sum_k \sum_j \left(g^{\eta}_{kt} + g^{\zeta}_{jkt} + g^{x}_{jt}\right)s_{jkt} \qquad (3)$$

where $s_{jkt} = \eta_{kt}\zeta_{jkt}x_{jt}/Y_t$. Collecting terms we can rewrite (3) as

$$g^{Y}_t = \sum_k g^{\eta}_{kt}s_{kt} + \sum_k \sum_j g^{\zeta}_{jkt}s_{jkt} + \sum_j g^{x}_{kt}s_{jt} \qquad (4)$$

where $s_{kt} = \sum_j s_{jkt}$, $s_{jt} = \sum_k s_{jkt}$. Equivalently, using $p_{jt} = \sum_k \eta_{kt}\zeta_{jkt}$,

$$g^{Y}_t = \sum_j g^{p}_{jt}s_{jt} + \sum_j g^{x}_{jt}s_{jt} \qquad (5)$$

Since data are available for discrete periods the continuous time forms are replaced by various discrete alternative index number formulations. For example, instead of (5) we could use the Tornqvist index

$$g_t^Y = \ln Y_t - \ln Y_{t-1} = \sum_j \tfrac{1}{2}(s_{jt} + s_{jt-1})(\ln x_{jt} - \ln x_{jt-1}) \quad (6)$$

or a Laspeyre index

$$g_t^Y + 1 = \sum_j (p_{jt-1}x_{jt}) / \sum_j (p_{jt-1}x_{jt-1}) = \sum_j s_{jt-1}(x_{jt} / x_{jt-1}) \quad (7)$$

With data on all the variables in (1) over time the choice of index number form is of secondary importance. However, as we will see later, the index form does matter when there is missing data.

The social cost of the inputs *z* used to produce NHS output is

$$Z_t = \sum_i w_{it} z_{it} \quad (8)$$

where w_{it} is the marginal social cost of input *i* in period *t*. The rate of growth of *Z* can be decomposed as a weighted sum of the growth rates of the volume of inputs and their marginal costs

$$g_t^Z = \sum_i (g_{it}^w + g_{it}^z) r_{it} \quad (9)$$

where $r_{it} = w_{it}z_{it} / \sum_i w_{it-1}z_{it-1}$. Again we use index number forms in practice because the data relate to discrete periods.

The discussion above has been couched in terms of overall NHS output and productivity but applies also to the measurement of output and productivity of lower level units from general practices to hospital Trusts.

The standard approach for measurement of private sector productivity is use market prices to aggregate different types of output and input prices to aggregate inputs. In the absence of market prices for outputs the current NHS output index uses unit costs to aggregate a subset of NHS activities.

3 Some problems in measuring productivity

3.1 Non standard outputs

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

3.1.1 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: eg healthy eating or anti-smoking campaigns. There are obvious problems in quantifying these outputs and their outcomes.

3.1.2 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.

3.1.3 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.

3.1.4 Training

The NHS invests in training its staff (for example via the NHS University). But not all staff whose human capital has thereby been increased continue to work in the NHS. Should we count the value of training for staff who leave as the increase in the present value of the income stream which they could earn in the private sector due to their training.

3.1.5 Research

The NHS funds a range of research. Given that information is a classic public good how do we value it?

3.2 Activities or outputs as the unit of analysis

3.2.1 Activities: institutional approach

NHS productivity measures have been based upon estimates of the number of certain 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, as such, 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, 2002).

The major disadvantage is that most patient cases pass through several institutional settings. Thus, for example, a patient who has a hip replacement will typically have been seen in general practice, in an outpatient department, treated as a 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 the valuation of output of separate sectors contributing to joint production across sectors problematic.

Current routine administrative data systems do not enable us to track the resource use associated with individual patients as they move along care pathways between 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

prescriptions or for what condition.

3.2.2 Outputs: patient-centred or disease-based approach

The bulk of NHS elementary 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 et al., 2001, Cutler and Huckman, 2003, Shapiro and Shapiro, 2001) 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 is 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 medium term the lack of properly linked routine data suggest that the measurement of NHS output will in fact be predominantly based on the measurement of activities.

3.3 Activity data availability

3.3.1 Coverage.

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[12]. 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 is less readily available means that estimates of NHS productivity growth may be biased downwards (Yuen and Towse, 2003).

3.3.2 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).

3.4 Valuation of outputs

3.4.1 Market prices

Under certain conditions the market prices for goods and services measure their marginal social value and hence can be used to aggregate them to construct measures of the growth rate of output. One possible method of valuing NHS output would be to use prices from the private sector. Some NHS activities have close matches in the private sector (non-emergency ambulance transport is similar to a taxi service for example). But the use of prices for private health care is problematic because of the presence of insurance so care is consumed beyond the point where its price equals the its marginal value to the patient. On the other hand 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. Moreover, the private sector does not produce the full range of NHS outputs. 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.

3.4.2 Unit costs

3.4.2.1 Theory

Current NHS practice, which follows the recommendation of the European Union, is to use production costs (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 cost as weights reflecting the marginal social value of outputs has the support, albeit reluctant, of Hicks (1940) but it rests on some very strong implicit assumptions about decision making in the public sector.

Imperfect agency. Suppose that there is a single taxpayer-patient with preferences over income y and two NHS outputs x_1, x_2 representable by the utility function

$$V \mid V^1(x_1, x_2)\, 2\, V^2 \mid y\, 4\, C \mid x_1, x_2 \,00 \qquad (10)$$

where C is the cost of NHS output which is covered by a non-distorting tax. See Figure 3. The taxpayer-patient will choose an output mix at a satisfying $V_j^1 \mid V_y^2 C_j$, 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.

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 satisfying $G \mid G(V), G\Re(V) \} 0$. Then the agent 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

$$G^b \mid V^1(x_1, x_2) 2\, kV^2(y 4\, C(x_1, x_2)),\ 0 < k < 1 \qquad (11)$$

who thus respects the tax-payer'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 3 coincide. Output will be at *b* where $V_j^1 / V_y^2 \mid kC_j$ 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-patients 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

$$G^c \mid G^{c1}(x_1, x_2) 2\, kV^2(y 4\, C(x_1, x_2)) \qquad (12)$$

which give rise to dashed indifference curves in Figure 3. The allocation will now be at a point like *c* where the marginal cost of x_j does not equal the taxpayer-patients marginal willingness to pay for it and $V_1^1 / V_2^1 \prod C_1 / C_2$. 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.

Deadweight loss from taxation. We assumed above that the NHS was financed by non-distorting taxation which is in contradiction of 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 we believed that the political mechanism led to a level of public expenditure at the marginal valuation of public sector output was equal to its marginal social cost, the marginal cost of production is 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 x^s is chosen. Then the capacity is rationed amongst patients. Let $B(x,w)$ ($B_w < 0$, $B_{xx} < 0$, $B_{xw} < 0$) be the net social benefit from elective care when the waiting time is *w* and *x* 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 ($B_x(x,w) = 0$) and that all infra marginal patients have positive net benefits. The demand for care is $x = D(w)$ ($D' = 4B_{xw} / B_{xx} < 0$). In equilibrium $x^s = D(w)$ and the equilibrium waiting time is $w(x^s)$ $w' > 0$. The supply of care is chosen at the first stage to maximise $B(x,w)$ $4\,C(x^s)$ subject to the second stage rationing mechanism. Hence the supply of care chosen satisfies

$$d[B(D(w(x^s), w(x^s)) 4\, C(x^s)] / dx^s \;|\; B_x D' w' (x^s) 2\, B_w w' (x^s) 4\, C' (x^s) \qquad (13)$$

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

$$C' (x^s) \;|\; B_w (x, w) w' (x^s) \qquad (14)$$

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 ($B_x = 0$). But, despite the inefficient rationing mechanism (rationing by waiting imposes a deadweight loss on patients and does not generate any offsetting benefit to anyone else), 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 the change in outcomes associated with an output should be measured 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.

This 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.

3.4.2.2 Practice

The costs used in the CWAI are average costs whereas, the preceding discussion suggests that marginal costs should be used. It seems unlikely that average NHS costs equal marginal costs for all outputs. The requirement could be relaxed to an equal marginal cost to average cost ratio 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.

3.5 Outcomes

The alternative to valuing outputs via market prices or unit costs is to estimate the volume of the different types of outcomes (ζ_{jkt}) generated by the outputs and to value the outcomes (η_{kt}).

3.5.1 Quality Adjusted Life Years

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 health over time and a discounting rule for adding up the health changes occurring at different points in time. Most QALY data is on the *level* of QALYs for different conditions and disease groups (Tengs and Wallace, 2000) but we are investigating the availability of results from evaluation studies which have incorporated health related quality of life measures. EQ-5D has been widely adopted in clinical studies and population health surveys including the 1996 Health Survey for England. It forms the basis of an approach recently used by the World Health Organisation in international comparisons of health status (Mahapatra et al., 2002). It can be combined with data from actuarial life tables to compute the quality-adjusted life expectancy for the general population (Kind, 1994). These estimates have been used in modelling the potential QALY gains from smoking cessation and mental health programmes as well as acute interventions. The EuroQol Group maintains a register of clinical studies and we hope to be able to use it to obtain EQ-5D scores pre- and post-intervention for a number of these applications.

Such studies provide estimates of QALY gains from an intervention at a particular date. But the QALY gains from treatment of patients with a given condition are likely to change over time because of technological progress and because of changing treatment thresholds. Cataract surgery is an obvious example. Thus we suspect that whilst evaluation studies will be a useful limited first step a more systematic routine collection of QALY change data will be required. That it is possible to collect QAly data on patients routinely has been demonstrated by the Health Outcomes Data Repository (www.cardiff-research-consortium.co.uk/hodar) and BUPA. HODAR has produced QALY estimates by ICD4 code by sampling patients after treatment. However numbers in different codes are small at the moment and the data is a single snapshot, not a record of changes due to treatment. BUPA has been administering SF36 routinely to patients before and after treatment.

3.5.2 Valuing outcomes

QALY gains. The National Co-ordinating Centre for Research Methodology (www.publichealth.bham.ac.uk/nccrm) has funded a programme of research to 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.

The value of reduced waiting time could be estimated by multiplying the value of a QALY times the change in QALYs generated by reduced waits. However, this is likely to underestimate the value of shorter waiting times. An alternative approach is to treat reduced waiting time as an outcome valued in its own right. The demand for health care is a demand for a bundle of characteristics of which waiting time, amenities and health outcomes are three important components. Indirect evidence from the market for private health care would support this approach. Health outcomes from treatment are unlikely to be different if a patient (who can afford it) chooses private or NHS treatment but the waiting time is significantly lower and the amenities higher in the private sector. This is what makes the private sector attractive (Beslely, Hall and Preston, 1999). Hence the prices for private treatment of some elective conditions may provide valuations of reduced waiting times.

The same approach could be taken to valuing improved amenities. 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 but are ignored in the CWAI approach. In the private sector such costs would be reflected in the prices paid for goods.

There has been some work by economists on the valuation of waiting times (Cullis, Jones and Propper, 2000). Imputed values for the standards of space and amenity of the physical infrastructure might be estimated from investment in maintaining and upgrading buildings in other sectors more responsive than the NHS to consumer demand. Obvious examples include hotels, office blocks, solicitor's offices, airports and private sector hospitals.

Finally, stated preference techniques can be used to derive monetary valuations of different outcomes (willingness to pay) or the marginal valuation of one outcome in terms of other outcomes.

3.6 Conclusion

Our tentative conclusion is that for the bulk of NHS activity there is currently no feasible alternative to using costs to weight activity in constructing an output index.

However this does not mean that, even with existing data, it is not possible to improve on a pure cost weighted index and we present the results from some simple explorations below. One possibility, which space prevents our exploring here, is that NHS output could be measured by a weighted average of a pure CWAI index applied to part of NHS output for which there is no good outcome data and an index based on outcome data for the rest of NHS output. As outcome data improves more of NHS output would be entered in the latter and less in the former.

4 An 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. (The DH has recently revised its Cost Weighted Activity index by incorporating a range of information on waiting times, mortality rates and other outcome indicators, though not QALY gains.)

4.1 Method

We have so far identified a small number of studies which report *changes* in health status as measured by EQ-5D (Bosch, 1999, Brilstra, 2004, Eefting, 2003, Garry, 2004, Lloyd, (in press), Manca, 2003, Nathoe, 2003, Ostendorf, 2003, Prinssen, 2004, Sculpher, 2004, Tangelder, 1999) and identified the HRG that would best describe the patients included in the study. This generated a set of nine HRGs, listed in Table 1, 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:

∉# For each of HRG, the number of finished consultant episodes (x_{it}), median waiting time (w_{it}) and mortality rate (m_{it}) were extracted from the Hospital Episode Statistics for each year from 1996/97 to 2002/03, where i indexes the HRG and t index the year. Both w and m improve on the estimates of changes in waiting times and mortality rates in the current CWAI by being specific to each HRG, rather than being aggregate measures, thereby being 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_{it}) were derived from the Reference Cost Schedules from 1997/98 to 2002/03, and are calculated 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_i), as measured by EQ-5D.

Time series data on FCEs, weighted reference costs, median waiting times and mortality rates for each HRG are in Figure 4

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_{it}).
2. A Laspeyre QALY gainweighted index, QWAI, in which cost-weights are substituted for the change in EQ-5D score (q_i).

These indices were then adjusted to incorporate changes in either median weighting times (w_{it}) or mortality (m_{it}). Both sets of indices were calculated from a base period

of 1997/98, when the Reference Costs were first published. The construction of the indices is shown in Table 2.

4.2 Results

Figures 5 and 6 track the various indices over the periods for which data are available. The indices suggest overall productivity improvements over the period for this basket of HRGs, with the extent of improvement dependent on which index is used and how much weight is placed on each of the components comprising the index.

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. The relative cost and QALY weights for the HRGs examined here one would expect productivity estimates to be quite different. Table 1 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 weighted in the QALY-based index.

However, for this set of HRGs, the overall effect of using these different weights is not substantial, with the CWAI and QWAI indices tracking each other quite closely, as shown in Figure 5. Of course, it cannot be inferred from this that a similar lack of overall sensitivity would apply to the valuation of other activities!

4.3 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. Moreover, 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 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 7 do not show much temporal variation but of course these are crude indicators of patient mix. We will investigating other possible indicators, such as more elaborate measures of demographic mix and the use of attributed socio-economic indicators from the 2001 Census.

The use of published studies to estimate changes in health status is not without precedent (Berndtand Busch and Frank, 2001) and may be combined with or validated by expert opinion (Berndtand Birand Buschand Frank and Normand, 2002). Nevertheless, there are some key 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 above, 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.

In order to measure accurately the change in health status as a result of treatment for the purposes of a productivity index, consideration should be given to the following:

⊄# Administering a generic instrument of health outcomes, such as the EQ-5D, both prior to treatment and post-treatment. The appropriate follow-up time may vary by activity.

⊄# Ensuring a representative sample of NHS patients and coverage across all activities comprising the index classification. The number sampled per activity will depend on the variability of health outcome experience among patients who receive the service.

⊄# Periodic re-sampling, to capture changes in health outcomes that may result from technological innovations in the provision of each activity.

This may seem an ambitious data collection exercise but experience suggests it to be feasible for the health service to collect such data routinely from a sample of patients. BUPA, for example, have collected SF-36 routinely from insured patients. In Wales, the Health Outcomes Data Repository (www.cardiff-research-consortium.co.uk/hodar) has administered EQ-5D to NHS patients, although much of the data collected have been post-treatment only.

References

Besley T, Hall J, Preston I. The demand for private health insurance: do waiting lists matter? *Journal of Public Economics* 1999;72:155-181.

Berndt, E.R., Bir, A., Busch, S.H., Frank, R.G., 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., Frank, R.G., 2001. Treatment price indexes for acute phase major depression, in: Cutler, D.M. Berndt, E.R., eds., Medical Care Output and Productivity, (University of Chicago Press, Chicago) 305-362.

Bosch, j.L., van der Graaf, Y., 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., 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,

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

Cullis JG, Jones PR, Propper C. Waiting times and medical treatment: analysis and policies. In: Culyer AJ, Newhouse JP, editors. *Handbook of Health Economics.* Amsterdam: Elsevier, 2000.

Dawson, D., Gravelle, H., Kind, P., O'Mahony, M., Street, A., and Weale, M. Developing new approaches to measuring NHS outputs and productivity. Project Protocol. March 2004.

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

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, 2870-2876.

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.

Hicks, J.R., 1940, The valuation of social income. Economica., Economica 7, 105-124.

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

Lloyd, (in press),

Mahapatra, M., Salomon, J.A., Nanda, L., 2002. Measuring health values in developing countries - results from a community survey in Andra Pradesh, in: Murray, C.J., Salomon, J., Mathers, C.D. Lopez, A.D., eds., Summary Measures of Population Health, (World Health Organization, Geneva)

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.

Nathoe, H.M., van Dijk, D., Jansen, E.W.L., Suyker, W.J.l., 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, 394-402.

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

O'Mahony, M., DeBoer, W., 2002, Britain's relative productivity performance: has anything changed?, National Institute Economic Review January, 72-82.

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, 302-309.

Prinssen, M., Buskens, E. Blankensteijn, J.D. and on behalf of the DREAM trial parteicipants, 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.

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., 2001. Measuring the value of cataract surgery, in: Cutler, D.M. Berndt, E.R., eds., Medical Care Output and Productivity, (University of Chicago Press, Chicago) 305-362.

Tangelder, M.J.D., mcDoennel, J., Van Bussbach, J. J., Buskens, E., Algra, A., Lawson, J. A., Eikelboom, B. C.,, 1999, Quality of life after infrainguinal bypass grafting surgery, Journal of Vascular Surgery 29,

Tengs, T.O., Wallace, A., 2000, One thosand health-related quality-of-life estimates, Medical Care 58, 583-637.

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

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

Yuen, P., 2002, OHE Compendium of Health Statistics, 13th edition (Office of Health Economics, London).

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

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, Casey Quinn, Sue Macran and Adriana Castelli for assistance with data and to Rowena Jacobs, Alan Williams, Alan Maynard, Lucy Stokes, Peter Smith, Luigi Siciliani, Karen Bloor, Philip Stevens, Jack Triplett for helpful discussions.

Table 1: 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	79.98%
E15	Percutaneous Transluminal Coronary Angioplasty (PTCA)	32,982	0.16	£2,959	66	0.72%	62	73.28%
H01	Bilateral Primary Hip Replacement	316	0.42	£5,791	254.5	1.27%	62	44.94%
H02	Primary Hip Replacement	42,354	0.42	£4,672	235	0.85%	68	39.14%
M07	Upper Genital Tract Major Procedures	63,925	0.18	£2,267	63	0.12%	45	0.03%
M13	Non-Surgical Treatment of Genital Prolapse or Incontinence	3,093	0.04	£592	79	0.03%	57	0.00%
M14	Non-Surgical Treatment of Fibroids, Menstrual Disorders, or Endometriosis	19,304	0.04	£543	33	0.10%	34	0.08%
Q05	Extracranial or Upper Limb Arterial Surgery	4,029	0.04	£2,412	28	1.27%	67	61.68%
Q12	Therapeutic Endovascular Procedures	14,749	0.10	£1,729	28	1.65%	66	58.01%

*Proportion discharged dead

Table 2: 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	$\frac{\sum_j x_{jt} w_{j0} c_{j0}}{\sum_j x_{j0} w_{jt} c_{j0}}$	$\frac{\sum_j x_{jt} w_{j0} q_j}{\sum_j x_{j0} w_{jt} q_j}$
Adjusted by mortality rate	$\frac{\sum_j x_{jt} m_{j0} c_{j0}}{\sum_j x_{j0} m_{jt} c_{j0}}$	$\frac{\sum_j x_{jt} m_{j0} q_j}{\sum_j x_{j0} m_{jt} q_j}$

Figure 1. Productivity, efficiency and welfare

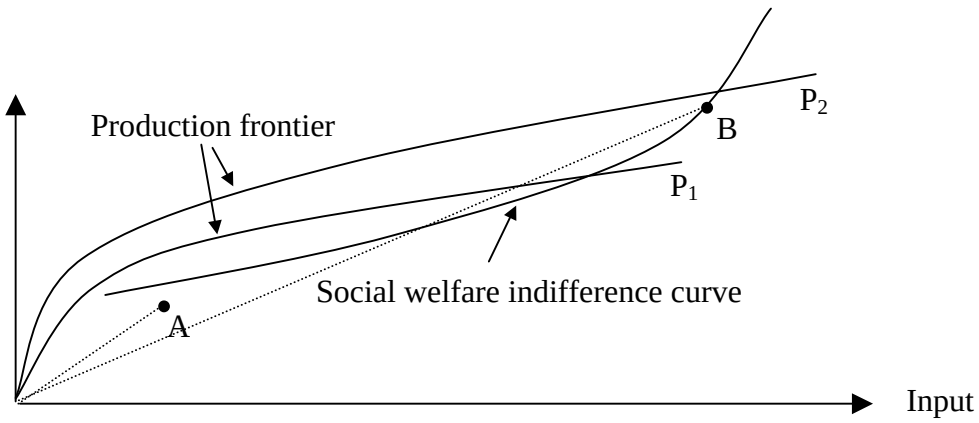


Figure 2. HCHS productivity and input growth rates

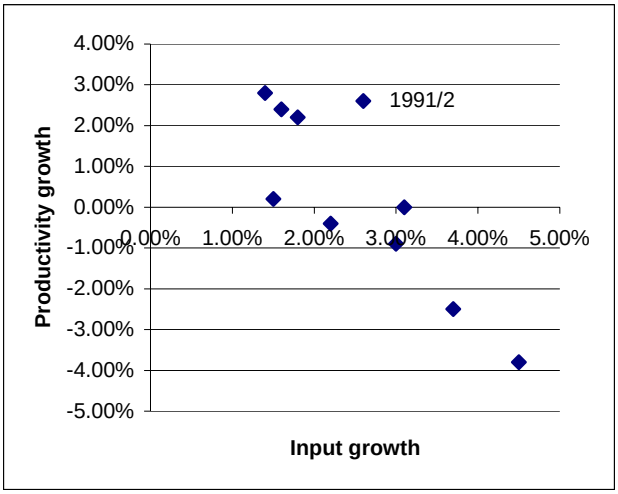


Figure 3. Marginal costs as marginal values: imperfect agency

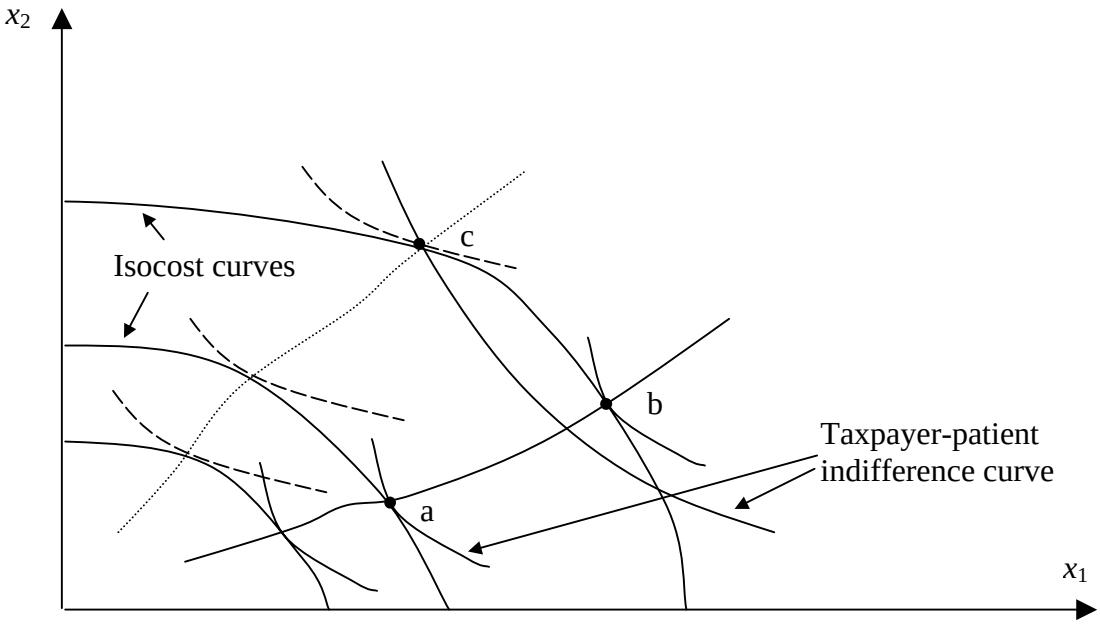
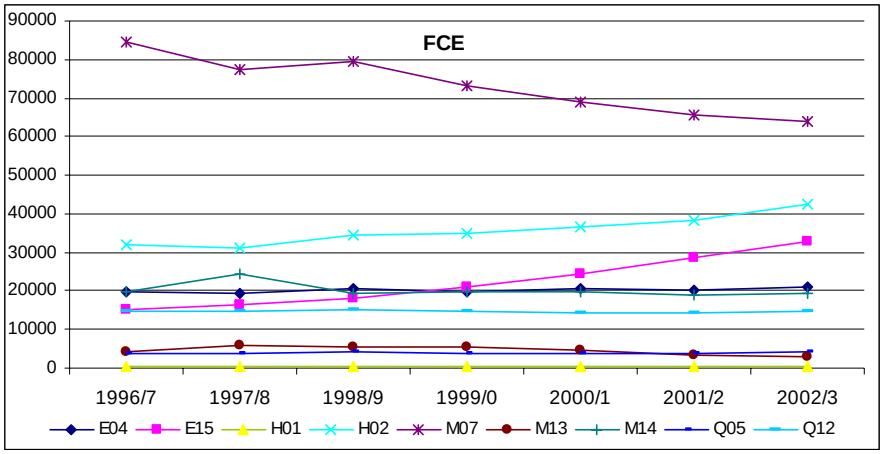
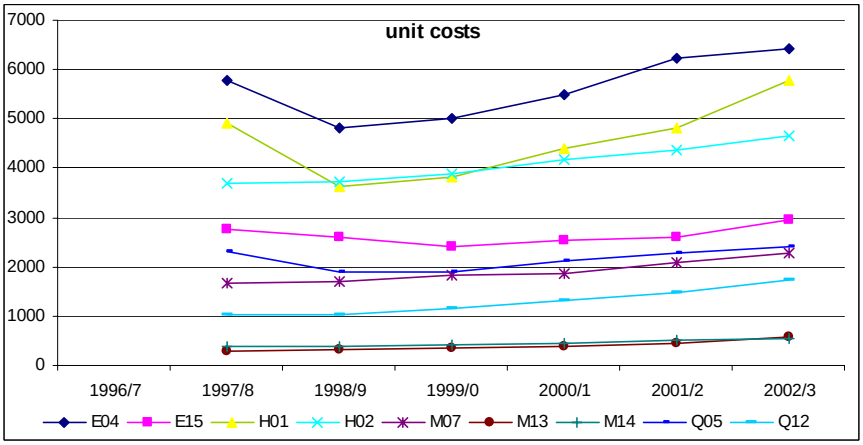
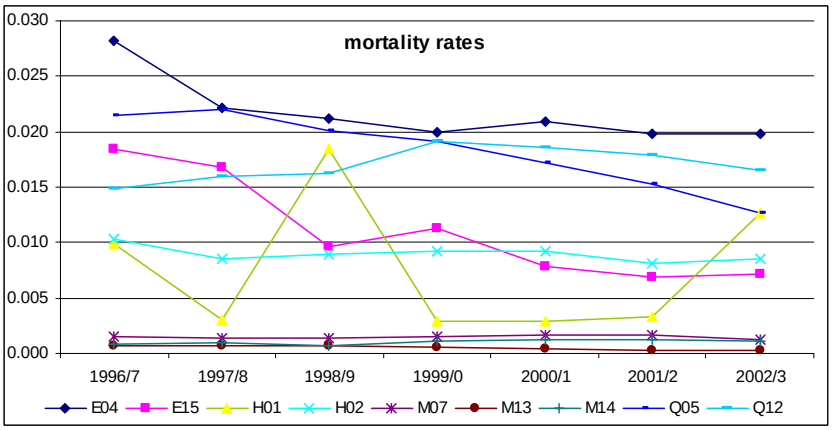


Figure 4. HRG Time series







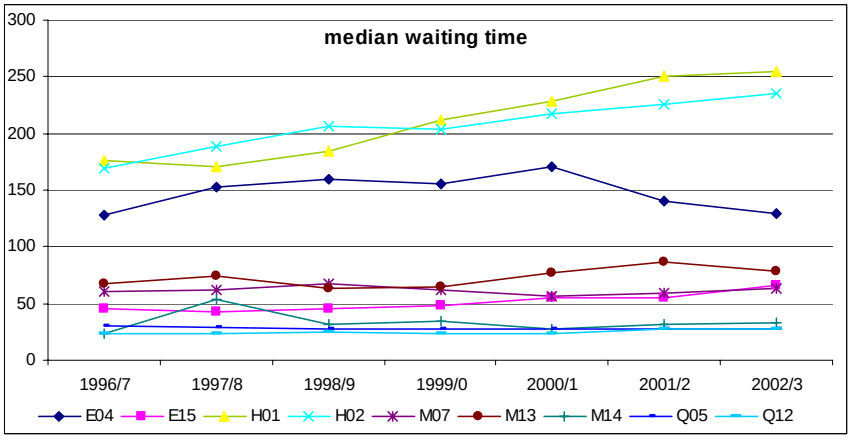


Figure 5: Comparison of based weighted CWAI and QWAI

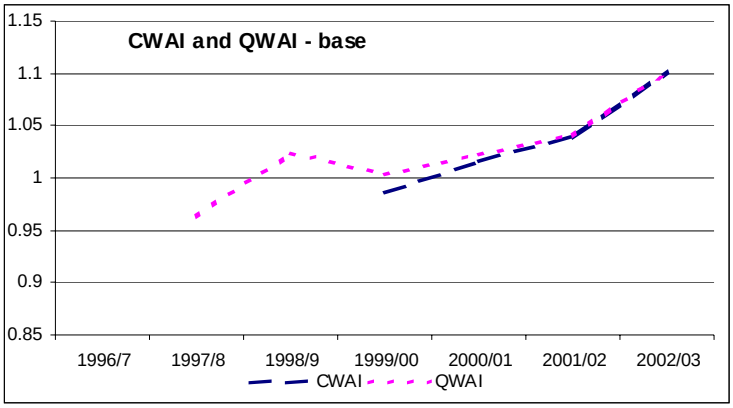
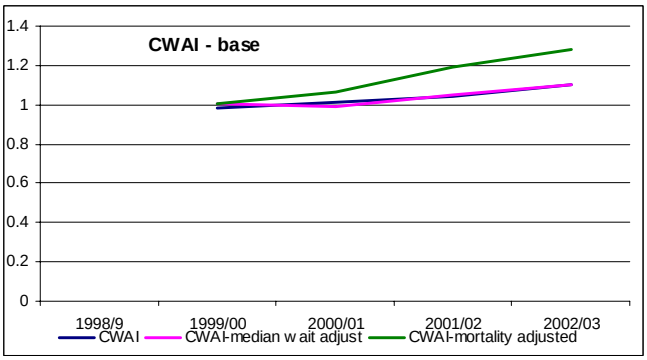


Figure 6. Adjusted CWAI and QWAI indices



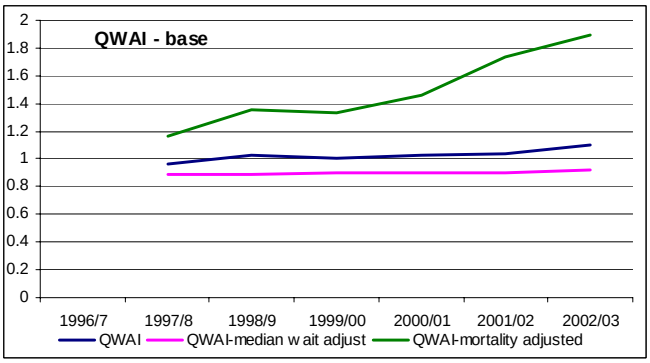


Figure 7. Patient mix