THE UNIVERSITY of York

CENTRE FOR HEALTH ECONOMICS

Developing New Approaches to Measuring NHS Outputs and Productivity
Second Interim Report
Data for Productivity Estimates

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

CHE Technical Paper Series 34

CENTRE FOR HEALTH ECONOMICS TECHNICAL PAPER SERIES

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

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

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

November 2004

Developing new approaches to measuring NHS outputs and productivity

Second Interim Report: Data for Productivity Estimates

November 2004

Lead Investigators:

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

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

With the assistance of:

Adriana Castelli[*], Sue Hennessy*, Rowena Jacobs[*], Sue Macran[*],

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

CHE Technical Paper Series No. 34

---

[*]Centre for Health Economics, University of York. **National Primary Care Research and Development Centre, Centre for Health Economics, University of York. ***National Institute for Economic and Social Research. Contact points: methodology – hg8@york.ac.uk; mweale@niesr.ac.uk; outcomes: ads@york.ac.uk; inputs: m.omahony@niesr.ac.uk

Executive Summary

1. This project has a long-term objective: to identify data and methods that can be employed by the Department of Health over future years to measure output and productivity in the NHS. In our First Interim Report we outlined a preferred methodology for measuring NHS output and productivity. In September 2004 we submitted a memorandum on data that the DH needs to collect if the long-term objective is to be realised. In this report we have focused on identifying data that can be used to measure activity in primary care, quality in secondary care, and volume measures of inputs and how they can be used to improve measurement of NHS productivity growth.

2. Measurement of NHS output requires quantification of the attributes of output valued by individuals. We have identified six key outcomes of NHS outputs which it may be possible to measure:
    - Health outcomes
    - Readmission rates
    - Waiting times
    - Choice and certainty of date of treatment
    - Patient satisfaction
    - Environment

3. Our approach has two important advantages over current methods for measuring NHS output. First, it permits measurement of changes in the quality of NHS output. For any given level of activity, if quality is increasing over time, the value of the NHS will be seen to be increasing. This change in quality will be captured by the change in the five outcomes we will attempt to measure. Second, our preferred methodology is consistent with that used to measure output and productivity in other sectors of the economy. Adoption of this approach could help to reduce the anomalies created by traditional methods that measure private sector output by reference to value to consumers but public sector output by reference to the costs of production.

4. The DH requires a productivity measure that can be used to track year to year changes in the volume of NHS outputs and inputs. Our preferred methodology for approaching this objective is that of total factor productivity growth: the difference between the rates of growth of indices of the value of NHS outputs and inputs. It represents the improvements in the use of existing resources through technical change or more efficient deployment of inputs but may also capture measurement errors and short term changes that may be due to some combination of chance or luck. Annual estimates of productivity growth should be treated with caution, given the existence of noise and lags in adjustments, so our results will concentrate on the implications of our methods for the estimation of changes in output and expenditure over a longer period.

5. Our First Interim Report reviewed availability of activity data at the secondary level. In this report we focus on data that can be used to improve measures of activity in primary care. Currently two measures of activity in general practice are included in the revised NHS outputs index: consultations and prescriptions. The information on consultations is derived from a general population survey (the General Household Survey). We have recommended (Appendix 1) that estimates of consultation activity would be better derived from databases extracted from electronic record systems in samples of general practices as these have a wider coverage of patients and contain more detail on consultations. After discussions with the Department of Health and the Office of National Statistics we propose to use the QRESEARCH database (which has data from 468 practices with about 7 million patients) in phase three to:

- estimate general practice consultation rates in different settings (surgery, home, phone) by different professionals (GP, nurse) over a period long enough to establish trends;
- investigate whether it is possible to use the rich information in the database to provide a finer classification of consultations and patient journeys using READ codes, prescribing and referrals data, and whether it is possible to measure trends in quality using indicators similar to those in the new GP contract.

We will also compare the estimates of consultation rates with those from the General Household Survey, the Health Survey for England, and the British Household Panel Survey.

6. While the DH collects a great deal of activity data, little information has been available on the quality of care. For this part of the project we have concentrated on identifying data that can be used to measure quality.

7. We have examined existing data on health outcomes. Continuous measurement of quality change over time requires the routine collection of health outcome data for all major interventions. No English Trust has collected such data. However, a large Welsh Trust has operated a continuous health status survey since June 2002 (Health Outcomes Data Repository [HODaR] database). This database has been made available to the research team and we report the results of our analysis of the data. Our conclusion is that the data are not suitable for the purpose of constructing an output index. The survey was not designed to produce information on before and after treatment outcomes. However, it does provide an indication of the variation in health outcome scores for particular conditions and this information might be used to assess the sample size requirements for the routine collection of observational outcome data in the NHS.

8. Two other observational databases have been identified that may be of use in preliminary estimates of quality change for a small subset of secondary care activity. BUPA has been collecting data on pre and post treatment health states for the last six years from seventy private UK hospitals. The database currently consists of 90,000 patient episodes for twenty high volume procedures. This database has recently been made available to the research team. The primary aim of our analysis will be to identify underlying general trends in changes to health status for particular procedures as a means of triangulating other methods of estimating of quality improvements.

9. York District Hospital has been collecting pre and post treatment data on health state since 2001 for two high volume orthopaedic procedures. The

research team will have access to this database. These data will be used to construct estimates of productivity growth for these procedures to triangulate estimates from other methods.

10. The experience of the Trust should also throw light on the costs and administrative issues relevant to routine collection of outcome data in NHS Trusts.

11. In the absence of data on health outcomes for the full range of NHS secondary activities we investigated the usefulness of evidence from clinical trials for measuring quality change. We report results from a survey of clinical studies employing a generic health status instrument, EQ-5D. Thirty studies were identified that reported UK data on pre and post treatment outcomes. We present summaries of the procedures and key outcome data from these research papers. We comment on the major problems with use of published clinical trial data to measure changes in the quality of NHS output. The most important are the small proportion of secondary care activity covered by such studies and the fact that they at best can be used to measure quality at one point in time, rather than changes in quality over time. In phase 3 of the project, we will explore the use of some of this data to triangulate other methods of estimating changes in quality

12. The only health outcome data routinely available across secondary care is for hospital mortalilty. Less than 3% of all NHS hospital episodes end with death of the patient. This is why we attach so much importance to obtaining data on the quality of outcomes for the remaining 97% of episodes where the patient may be expected to benefit from treatment and remains alive. We have made suggestions on how this might be done in our data requirements note to the DH (Appendix 1). However, at present any attempt to measure quality change across all secondary activity will have to use information on post treatment mortality. In this report we reviewed the data and methods available for measuring changes in quality adjusted life expectancy for NHS patients. Preliminary estimates of "quality adjusted lives saved" are presented to illustrate possible uses of the data. We will use and extend these methods in

phase three of the project.

13. We report on data available to measure outcomes other than health outcomes. We expect that in the short term the most important of these non health outcomes will be changes in waiting time. Ideally we want to measure changes in waiting time from GP referral to admission for treatment by procedure (or HRG). With present DH data this is not possible. Waiting time for a first outpatient appointment is only available at the specialty level. The DH has discontinued collection of waits for second or more outpatient appointments. In the short term there is no choice but to use the more limited data on waiting time from the date a patient is added to the inpatient waiting list to the date of admission for treatment currently available from Hospital Episode Statistics (HES). The government has announced new waiting time targets that relate to the time between GP referral and admission. Assuming this target is to be monitored, the DH can be expected to change the requirements for reporting waiting times that will overcome the limitations of existing data.

14. Measuring improvement in giving patients choice of data and certainty of date of treatment relies on data from HES on the proportion of admissions recorded as elective-booked. Increased use of Independent Treatment Centres (ITCs) by the NHS may reduce the value of this measure of quality change. It is not clear what data ITCs will be required to submit to the DH. Most if not all of their activity should be "booked" but we also need to know their activity levels by HRG if it is to be aggregated with that of NHS Trusts. With existing data it will not be possible to monitor changes in cancelled operations. Data on cancellations *after* the patient has been admitted to hospital are not available.

15. One characteristic of NHS output gaining increasing importance is that of hospital cleanliness and the control of infection. We have reviewed the available data. Patient Action Teams (PEAT) inspect hospitals and report on the patient environment (cleanliness and tidiness) and food services. A number of different surveillance systems have been in place to monitor infections in English hospitals. These fall into three categories: wound

infection, overall measures of infection control and MRSA (hospital acquired infection). We report important problems in the coverage of this data. In phase 3 we will examine the potential for incorporating the limited existing data on cleanliness and infection rates in measures of quality change.

16. We have examined existing surveys of the NHS patient experience that have been carried out on behalf of the Healthcare Commission. Five national surveys have been undertaken. We do not have access to the raw information (individual patient responses) but do have summary statistics on the proportion of patients responding to each question in terms of the percent indicating the service in question was poor, very good, etc. We will be examining methods for incorporating this limited data into our quality measures.

17. A central problem in creating a single index of output is identification of weights that can be used to sum the diverse outputs of NHS activity. In theory the weights should reflect the marginal social value of the outputs. The problem is identifying data to estimate or approximate social value in the absence of market prices. We have explored three options: results of discrete choice experiments, estimates of marginal cost and international prices. Evidence from discrete choice experiments is scarce, studies are infrequent and it is difficult to extrapolate from sample to general population values. Valuations from discrete choice experiments are unlikely to be suitable for the generation of weights to be used in an index of output intended to routinely monitor changes in NHS output.

18. In our First Interim Report we argued that under certain assumptions, marginal costs may reflect marginal valuations of output. Current NHS practice is to use unit (average) costs derived from Reference Costs to weight diverse NHS activities. To test whether the use of relative marginal costs would make a difference to measures of output growth, we are attempting to estimate marginal costs and compare them to reference costs. However, there are not enough observations on the costs of individual Trusts to enable marginal costs to be estimated at HRG level and so any measure of output growth which attempts to apply different weights for activity in different HRGs will have in

the short term to continue to be based on average costs. Our estimates of marginal cost made a more aggregated level may however indicate if average costs are a poor approximation to marginal costs and hence whether attempts should be made to estimate marginal costs at HRG level by using the raw accounting data used to construct average cost estimates.

19. We were asked to look at whether international prices might provide plausible indicators of the relative marginal social value of different NHS outputs. Publicly accessible information is limited but we have been able to undertake a comparison of unit costs for England, Australia and Italy. We report the results of this preliminary study. We conclude that international prices, usually based on national costs of production rather than market prices, are unlikely to be of use in estimating the relative value of NHS outputs.

20. We will be exploring ways that the available data on quality change can be incorporated into a cost weighted activity index. This will include assessing how robust productivity estimates are to alternative approaches to combining quality and outcome information with cost weighted activity.

21. We have examined data available for the measurement of inputs. We have identified data sources for the three broad input categories, labour, intermediate and capital. In all three cases combining a range of data sources should yield better estimates than available to now. Thus for labour input we will combine DH sources such as the NHS employment census and the NHS earnings survey with national sources such as the Labour Force Survey and the Annual Survey of Hours and Earnings. Intermediate and capital inputs will use data from Trust Financial Returns, Inland Revenue Inquiries, PACT, NHS estates and national sources on historical investment series. The measurement of inputs will proceed in collaboration with ONS.

Table of Contents

1 Introduction

This project has a long-term objective: to identify data and methods that can be employed by the Department of Health over future years to monitor output and productivity in the NHS. In our First Interim Report (Dawson *et al*., 2004a) we outlined a preferred methodology for measuring NHS output and productivity. In September 2004 we submitted a memorandum on data that the DH needs to collect if the long-term objective is to be realised (Appendix 1). In this report we focus on the data currently available and how other data, feasible to collect, can be used to improve measurement of NHS productivity.

1.1 Preferred methodology

We distinguish between *activities* (operations, GP consultations), *outputs* (courses of treatment which may require a bundle of activities) and *outcomes* (the characteristics of output which affect utility). Measurement of the value of NHS output requires quantification of the attributes of output valued by individuals. In general, the social value of NHS output over year t is:

$$Y_t = \sum_j \sum_k \beta_{kt} \alpha_{jkt} x_{jt} = \sum_j p_{jt} x_{jt} \tag{1}$$

where x_{jt} is the volume of output j over year t, α_{jkt} is the amount of outcome k generated by a unit of output j, β_{kt} is the average social value (expressed in monetary units) of a unit of outcome k in year t and p_{jt} is the average social monetary value (or price) of a unit of output. The price of output j is equivalently expressed in terms of the values of the vector of outputs it generates: $p_{jt} = \Sigma_k \beta_{kt} \alpha_{jkt}$.

We have identified five key outcomes of NHS output which it may be possible to measure:

- Health outcomes
- Waiting times

- Choice and certainty of date of treatment
- Patient satisfaction
- Environment

This approach has two important advantages over current methods for measuring NHS output. First, it permits measurement of changes in the quality of NHS output. For any given level of activity, if quality is increasing over time, the output of the NHS will be seen to be increasing. This change in quality will be captured by the change in the five outcomes we will attempt to measure. Second, our preferred methodology is consistent with that used to measure output and productivity in other sectors of the economy. Adoption of this approach could help to reduce the anomalies created by traditional methods that measure private sector output by reference to value to consumers but public sector output by reference to the costs of production.

There are many NHS activities where the lack of data means that in the short term outcomes cannot be measured. However, our approach to measuring output can be applied to subsets of NHS production such as treatment for a particular disease or condition where there is more information than for the rest of NHS activity.

Given the short term lack of data on outcomes for most NHS activities, a "comprehensive" index of output will still have to employ cost weights to sum different activities. We will be exploring ways the available data on quality change can be incorporated into a cost weighted index of output.

Improved measures of NHS output are likely to be of use to clinicians, Trusts and those responsible for monitoring the service. However, the Department of Health is particularly interested in two applications: estimates of value for money and estimates of productivity. The literature reviewed in our First Interim Report revealed that it was rare for studies of value for money or productivity to focus on year to year changes. The noise observed in the data over short periods of time makes it difficult to identify underlying change. In addition there are significant lags in the impact of some forms of expenditure on output and outcomes.

The Department of Health interest is in a productivity measure that can be used to track year to year changes in the volume of NHS outputs and inputs. Our preferred methodology for approaching this objective is that of total factor productivity growth: the difference between the rates of growth of indices of the value of NHS outputs and inputs. Annual estimates of productivity growth should be treated with caution, given the existence of noise and lags in adjustments, so our results will the implications of our methods for the estimation of changes in output and expenditure over a longer period.

1.2 Structure of the report

Section 2 sets out progress to date in identifying data suitable for measuring NHS outputs. Given our interest in measuring quality change, we focus on the data that may be used to identify trends in the quality of NHS outputs. Section 3 summarises data available for measuring quality adjusted inputs. Section 4 contains our conclusions to date and implications of this review of data availability for phase 3 of the project which will empirically investigate methods of estimating productivity growth.

2 Measurement of activity

There are comprehensive data on activity at the hospital level. They are generally available and currently used by the DH and ONS to estimate NHS output. In our First Interim Report (Dawson *et al*., 2004a, Appendix E) we listed the fields of main interest in measuring hospital and community and health service activity available in the Hospital Episode Statistics and reference cost returns. The two main problems with routine NHS activity data are (1) lack of information on activity in primary care and (2) lack of a patient identifier that permits tracking patients across service boundaries. If, as expected, there will be an increasing shift of treatment from hospital to GP surgery or community care, these data deficiencies will become more serious for attempts to measure NHS output. Our memorandum to the DH on data requirements (Appendix 1) addressed some of these issues.

Here we focus on data that may be used to improve analysis of activity in primary care.

2.1 General practice

Currently two measures of activity in general practice are included in the revised NHS outputs index: consultations and prescriptions. In previous years the CWAI did not include consultations and prescriptions but did include such activities as cervical screening tests, and visits by district nurses and health visitors.

Ideally NHS productivity measures should be based on numbers of patient journeys of different types where journeys are likely to involve both primary and secondary care. In the absence of routine record linkage such measures are not currently feasible but it would still be worthwhile getting a finer breakdown of GP consultations to allow for the changing mix of providers and for the changing mix of types of consultations.

Estimates of the volume of consultations have to be obtained from samples of patients or practices. The data have to be capable of providing regular estimates of consultation activity during each year (and ONS require quarterly estimates for their productivity measures) broken down by

- provider (GP, nurse – any other providers in the practice – e.g. physiotherapy)
- age, gender of patients to enable the data to be grossed up to yield national estimates. (Ideally we would want to have richer data on patients on characteristics known to affect consultations which are also measured in the 2001 Census such as ethnicity, and education.)
- diagnostic category
- result of consultation – prescription, referral, monitoring of condition, advice etc

2.1.1 Consultations: GP electronic record systems

Most general practices have electronic patient record systems which can be used to record details of consultations. There are a number of databases which collect information from these systems from samples of practices including the General Practice Research Database, QRESEARCH (University of Nottingham), Royal College

of General Practitioners Weekly Returns, and International Medical Statistics. The largest seems to be the QRESEARCH database which already has a contract with the DH to providing data in connection with the new GP contract (see below).

QRESEARCH (QR) currently has data from the record systems of 468 practices whose geographical distribution and patient characteristics suggest are a reasonably representative national sample. (www.qresearch.org; Hippisley-Cox, 2004). It is possible to calculate consultation rates for different age and gender groups, by provider (GP, practice nurse), and location (surgery, home, phone, elsewhere). The information is available from 1988 though the small number of practices in the early years suggests that estimation of trends in consultation rates should be limited to the period from the mid 1990s onward (in 1995 there were 183 practices with a total registered population of 1.27M).

We feel that the QRESEARCH database is the most promising source of information on consultations. Previous analyses of the information in the database (Hippisley-Cox, 2004) shows that it is possible to use it to estimate consultation rates. The practices and populations seem nationally representative but a number of detailed questions require to be addressed. It is for example possible that practices recording of information improves after they join the database so that estimated consultation rates would be affected by the rate of inflow of new practices. We will test for these and similar effects and examine their implications for the way in which the data can be used to generate consultation rate estimates. For example, if recording of information improves for newly joining practices should there be a lag before their data are used to compute national estimates or is some other type of adjustment preferable.

After discussions with the Department of Health, the Office of National Statistics, and the shadow Information Authority (which is likely to take over the existing QRESEARCH contract in April 2005) we propose to use the QRESEARCH database to
(a) estimate general practice consultation rates in different settings (surgery, home, phone) by different professionals (GP, nurse) over a period long enough to establish trends;
(b) investigate whether it is possible to use the rich information in the database to provide a finer classification of consultations and patient journeys using READ codes,

prescribing and referrals data, and whether it is possible to measure trends in quality using indicators similar to those in the GP contract (see below).

The analysis will take place in two stages. The first stage will be for the investigation of the representativeness of the database and the estimation of numbers of consultations. The DH and ONS would like to have estimates of numbers of consultations for the period from 1996/7 by April 30 2005 for the next revision of DH activity index and incorporation in the ONS estimates of productivity. This suggests that the stage 1 data extract should take place as early as possible in January 2005 to allow for investigation of representativeness etc. We have prepared a draft data extract specification for this stage and are in the process of refining in discussion with QRESEARCH.

The second stage data extraction will be of a finer set of classifications of consultations and would require fields additional to those extracted for stage 1. The final project report is due August 31 2005 so that the stage 2 data extract would be required in May 2005. We will produce a second detailed functional specification for the stage 2 extract in the light of experience with examination of the data in the stage 1 extract.

2.1.2 Consultations: survey data

Three nationally representative surveys have information on consultations.

General Household Survey. Estimates of consultation activity can be derived from the consultations reported by respondents in around 9000 households in the General Household Survey (GHS). Respondents are asked about consultations in the previous 14 days. The estimate of the number of consultations per year is made by multiplying the number of reported consultations in the 14 days prior to interview by 26. Information is available on the location of the consultation (surgery, home, phone) and, after 2000, by provider (GP, practice nurse). From 1988 the GHS has collected data over a financial year (April to March). No allowance has been made for seasonal factors - the date of the consultation varies across respondents and has also varied between rounds of the GHS. There are two missing years as there was no GHS in 1997/1998 and 1999/2000. The GHS is the source of the estimates of consultations used in the current NHS productivity measure and also in the ONS productivity estimates.

British Household Panel Survey (*BHPS*). The BHPS was started in 1991 with 5500 households in Great Britain and currently contains 12 waves. Wave 1 (1991) had 5500 households in Great Britain. Boost samples of 1500 households in Wales and Scotland were added in 1999. 2000 households in Northern Ireland were added from 2001. The survey takes place in September to December each year. The BHPS has a question about GP use in the previous 12 months but does not distinguish place of consultation and does not enquire about other general practice consultations, for example with a nurse.

Health Survey for England (*HSE*). The HSE is an annual survey with around 15000 respondents (more in years with boost sample aimed at particular sections of the population). Fieldwork typically takes place from January to March/April of the following year. There were questions about consultations with a GP in the previous 2 weeks in 1993 and 1994 and then in every year from 1998. From 1998 there is information about the location (home, surgery, phone) and from 2001 onwards there have also been questions about consultations with practice nurses. There is some limited information on the purpose of the consultation (for CVD, mental problems) and on whether the consultation resulted in a prescription.

Although we believe that consultations are best estimated using data extracted from GP record systems rather than from self report of patients in population surveys we will investigate the three surveys as a source of consultation data in more detail in stage 3 of the project. We will compare the data on consultations from the GHS, the BHPS and the HSE for a common set of years and compare the determinants of consultation rates in the three surveys. The results will be compared with those from the QRESEARCH database derived from GP records.

2.1.3 GP cost weights

The PSSRU estimates the unit costs of GP and nurse consultations (http://www.pssru.ac.uk/pdf/uc2003/uc2003.pdf) using a variety of official and unofficial sources. Several of the estimates rest on self reported GP activity from the 1992/3 GP Workload Survey undertaken for the DDRB. There does not appear to be a

more recent survey of GP activity and we have previously suggested that DH should consider undertaking such a survey at regular intervals. We will use the PSSRU estimates where appropriate but will also examine the sensitivity of results to alternative assumptions.

2.1.4 Prescribing

The prescription activity measure in the recently revised NHS outputs index is derived from PPA data. The PPA data are collected in order to remunerate pharmacists (and dispensing GPs). It is therefore a comprehensive measure of prescriptions dispensed and can be disaggregated to product type if required. The data are reliable, comprehensive and readily available at national levels of aggregation. They have been used to construct a number of indicators of practice prescribing quality as well as quantity.

The usefulness of the data could be greatly improved and this would be relatively simple. The most obvious example is by improving the patient information on the prescription form. At the moment the only patient data on the form indicates if the patient is entitled to free prescriptions and on what grounds. This information has been used by the Prescribing Support Unit to produce the Low Income Scheme Index which measures the proportion of prescriptions which are dispensed without charge on grounds of low income. The LISI is the only direct variable measuring practice population socioeconomic status which relates directly to practice patients rather than being attributed from Census or Social Security data on the basis of patient postcode. Adding a field for diagnosis to the prescription form would greatly enhance the usefulness of routine prescribing data as a measure of prescribing quality. Adding gender and age fields would also improve the socioeconomic data and improve prescribing quality indicators. We have suggested (Dawson *et al*., 2004b, see Appendix 1) that the DH should consider adding these fields to the prescription form so that quality adjustments can more readily be made to the data on prescribing activity.

Most consultations result in a prescription. We will therefore consider in phase three whether including measures of both consultations and prescriptions in an output measure will lead to double counting and how the two activities should be combined in an overall output measure.

2.1.5 Other GP activity data

The new General Medical Services contract introduced in April 2004 provides financial rewards for practices related to 146 quality or performance indicators (Roland, 2004). The contract will greatly enrich the set of activities which are routinely measured and may provide a means of quality adjusting the measures of general practice output. The data will be centrally collected for the first time in 2004/5 in the Quality Management and Analysis System (QMAS) primarily for calculating practice remuneration. A separate database, the Quality, Prevalence and Indicator Database (QPID) is being developed to hold information from QMAS in a format which is suitable for analysis. The QPID database cannot be used for calculations of productivity growth until the 2005/6 data are available, which will be after the end of the current project. Estimates of productivity growth based on QPID data are also likely to be unreliable for the first few years of the new contract because of changes in data collection and coverage. It is possible that the QRESEARCH database may be a more reliable source of estimates of national trends in some of the quality indicators in the new GMS contract for these years. We not investigate this in phase three of the report given the time and resource constraints but we our experience with use of the QRESEARCH data to measure consultations may enable us to make some suggestions about how quality indicators might be developed and incorporated in the output measure in the longer term.

A sizeable (35%) proportion of practices have switched from GMS contracts to Primary Care Services (PMS) contracts which are locally negotiated with their Primary Care Trusts. PMS practices are expected to report how they perform against the quality indicators in the new GMS contract but it is as yet unclear how many will actually do so and how reliable such information will be. We have suggested that the DH should attempt to ensure that the QOF and targets data from PMS practices is comparable in quality and coverage to that for GMS practices.

2.2 NHS Direct, NHS Direct Online and Walk-In Centres

Accurate measurement of productivity improvements necessitates capturing technological improvements or changes in the nature of service delivery. Often these

changes will not be recognised by existing data coding practices, making it particularly challenging to capture their importance. NHS Direct, NHS Direct Online and Walk-In Centres are recent innovations in the provision of first contact advice, information and treatment. These initiatives are designed to fulfil a role not previously offered by the NHS and to act as a substitute provider of advice and health information (e.g. diverting activity from general practice). They are likely to reduce the costs to patients of such first contacts, leading both to an increase in primary care activities and to a change in the mix of activities in general practice. The organisations are expected to play an increasing role in the NHS over the coming years. The revised version of the NHS productivity measure uses some of the available information on these services and we will investigate whether there is value in extracting more information to estimate outputs.

2.2.1 Data identified

Aggregate data on use of NHS Direct and NHS Direct Online are available, as reported in the Chief Executive's Report to the NHS – Statistical Supplement (Department of Health, May 2004). The data differ slightly from those in the spreadsheets compiled (29 March 2004) by EOR for the revised NHS output index. The table below summarises their activities from the former source. The data indicate a gradual increase in activity since the two bodies were created.

Table 2.1 NHS Direct and NHS Direct online activities

Years	Calls received ('000s)	Visits to NHS Direct on line ('000s)
1998/99	110	n/a
1999/00	1,650	n/a
2000/01	3,420	1,500[1]
2001/02	5,213	2,028
2002/03	6,319	3,972
2003/04	6,411	6,542

Source: Chief Executive's Report to the NHS –Statistical Supplement (Department of Health, 2004a)
Note: NHS Direct on line was launched in December 1999, and figures for 2000/01 are an estimate.

Data on visits to Walk-In Centres are contained in the Chief Executive's Report to the DH, 2004, Statistical Supplement, which indicate an increase in visits since March 2000. These counts of activity differ slightly from the estimates contained in the

spreadsheets compiled (29 March 2004) by EOR for the revised NHS output index.

Table 2.2 Average activity in Walk-In Centres

Year	Number of sites	Monthly average number of visits to a site per day[1]	% change on same month in previous year
March 2000	4	30	
September 2000	30	55	
March 2001	39	69	132
September 2001	39	79	
March 2002	42	81	17
September 2002	42	92	
March 2003	42	94	16
September 2003	42	101	
March 2004	43	103	10

Source: Chief Executive's Report to the NHS –Statistical Supplement (Department of Health, 2004a)
Note: All figures are partially estimated.

We have recently received a detailed monthly breakdown of the activities for all 43 Walk-In Centres, covering the period April 2004 – September 2004. The data cover the following:

- Number of visits
- Day and time that the visit takes place
- Duration of visit
- Responsiveness of the service (e.g. how long before the patient was assessed)
- Patient age and gender
- Presenting condition
- Type of consultation
- Action taken, including whether or not the patient was referred to another health professional

2.2.2 Gaps in data availability

While detailed data are available for Walk-In Centres, little more can be ascertained from the aggregate data we have for NHS Direct and NHS Direct Online. It would be

useful to know more about the nature of their activities, especially as the bodies take on an expanding role. In order to measure the outputs of the services more accurately it would be helpful to have data on:

- the breakdown of enquires between the provision of health advice and information about the health service
- type of conditions people seek health advice about
- actions that are recommended as a result of the request

It appears that these data might be collected routinely for NHS Direct. The evaluation of NHS Direct conducted by the Medical Research Unit at the University of Sheffield was based on analysis of information collected routinely (Munro *et al*., 2000). The telephone service involves staff asking callers about the nature of their enquiry by using standardised triage software that, through a given set of questions, is used to identify and classify symptoms. This information is recorded electronically during the telephone conversation. It appears that information from the former source was available for the evaluation of NHS Direct. As yet, we have not been able to secure these data.

The website service for those who seek advice from a nurse involves self-completion of a detailed questionnaire on the nature of the symptoms and condition, as well as personal information (http://www.nhsdirect.nhs.uk). However, these data do not appear to be submitted to NHS Direct Online. It may be worth considering whether to extract such information in the future.

2.3 Other primary care services

The DH collects data on a wide range of other primary and community care activity such as ambulance services, community mental services, and cervical screening. The data are described in Appendices 2 and 3. Some, but not all, of these have been included in DH NHS output indices. In phase 3 we will examine alternative methods of incorporating them into estimates of productivity growth.

3 Output characteristics and quality

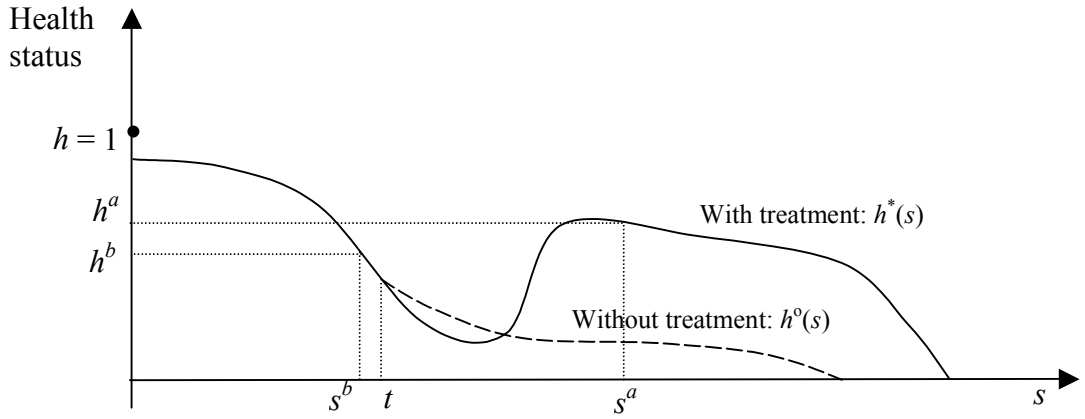
3.1 Health outcomes

The main aim of the health system is the improvement of the health of the population. This being so, it would seem reasonable that any measure of health system productivity should include measures of the effect of the system on health. Current practice in estimating NHS productivity growth ignores quality change and, in particular, the change in health gains from treatment. The challenges associated with measuring the effect of interventions are detailed and discussed in our First Interim Report.

The construction of a productivity index requires information about changes in health status attributable to interventions. Such information currently is not collected by the NHS. We have explored a number of approaches which can be used to estimate the health effects of treatment. These fall into four categories:

1. Quality adjusted post treatment mortality rates, which might provide estimates of the loss in quality adjusted life years due to mortality from treatment
2. Clinical trial data, from which it may be possible to extract with and without treatment estimates of health status for patients enrolled in clinical trials;
3. Observational data, from which before-and-after treatment estimates might be derived for patients in routine NHS practice;
4. Expert groups, which might be used to estimate the health status that patients would experience with and without treatment.

Figure 3.1



These approaches will yield different estimates of the impact of interventions on health status. The effect of health care is to change the time profile of stream of health status. In Figure 3.1 treatment takes place at time t. Without treatment the individual would have the time stream $h^o(s)$ and with treatment would get the time stream $h^*(s)$. Let $v^*(t)$ and $v^o(t)$ denote the discounted quality adjusted life years generated by the time streams with and without care discounted to date t. Thus

$$(2) \qquad v^*(t) = \sum_{s=t}^{s=t+L^*(t)} \delta^{s-t} h^*(s), \qquad v^o(t) = \sum_{s=t}^{s=t+L^o(t)} \delta^{s-t} h^o(s)$$

where δ is the discount factor applied to future health and $L^*(t)$ and $L^o(t)$ are the remaining lengths of life of the patient with and without treatment at date t. The outcome of treatment at date t is then $\Delta v(t)$ and this is what we would ideally wish to measure each year and for all types of care.

There is some debate about the appropriate choice of the discount δ to be applied to health status (QALYs) in evaluating interventions (see Gravelle and Smith (2001) for a discussion and references). Since the discount factor is unlikely to change markedly over time we do not propose to investigate it in phase 3. Instead we will consider whether estimates of productivity growth are robust to alternative assumptions about the time invariant value of δ.

For reasons discussed in our first report, it is difficult to derive estimates of $\Delta v(t)$,

primarily because without treatment profiles are not observed. Instead the type of data that are likely to be available comprise snapshots of the *level* of health status at a point before and a point after treatment. In terms of Figure 3.1, health status h^b before (at date s^b) and after h^a (date s^a) treatment would be measured. The snapshot heath status measures have the dimension of units of health. To translate $[h^a - h^b] = \Delta h$ into the same units as the required QALY change measure we must scale it some factor k which will vary with the particular treatment. We can estimate the true QALY gain $\Delta v(t)$ by $k[h^a - h^b] = k\Delta h$. In some cases we can interpret k as the period over which the health effects obtain. Although there will always exist some k such that $k\Delta h = \Delta v(t)$, the fact that we do not observe $\Delta v(t)$ means that k has to be based on judgement and possibly evidence from clinical trials.

An outcome measure based on the difference between post and pre-treatment snapshot health status measure is imperfect. It does not measure health with and without care but health before and after care. It also replaces each time profile with a single snapshot. For some treatments and conditions the effect of treatment is merely to slow down the rate of decline in health status, so that $\Delta h < 0$ even though $\Delta v > 0$. The fact that the scaling factor k is likely to be subject to considerable uncertainty means that estimates of the *level* of productivity based on $k\Delta h$ may be subject to considerable margins of error.

However, since the aim is to measure the rate of growth of productivity we are interested in whether the rate of growth of $k\Delta h$ is a reasonable approximation to the rate of growth of Δv. Changes in the time profiles $h^*(s)$, $h^o(s)$ arising from changes in the underlying technology or the delivery of care can alter both the before and after health status snapshots h^b, h^a and the scaling factor k. Since k will be based on judgements, whereas of Δh is estimated from observations, it seems sensible to be cautious in letting changes in judgements about k influence estimates of the rate of growth of $k\Delta h$.

The default assumption should be that the scaling factor is constant over time and should only be changed when there is good evidence that there would otherwise be serious inaccuracy in estimating productivity growth. With this convention the choice of scaling factor affects the level of productivity but makes no difference to the estimated

rate of growth of productivity. The important issue is then how well the rate of change in the snapshots approximates the rate of change in the areas under the two time profiles.

Both the level of health before treatment h^b and the health of treated patients if not treated depend on the patient population selected for treatment and on the general health of the population. It is not unreasonable to suggest that the rates of change of h^b and the discounted value v^o of the no treatment health profile $h^o(s)$ over time will be similar. Both the snapshot level of health after treatment h^a and the discounted value v^* of the time profile h^* will be measured on the same population and hence are affected by the same factors including any technological change.

Hence, despite the imperfections of the difference between snapshots of post and pre treatment heath status $k\Delta h$ for calculating the *level* of productivity, we suggest that the rate of change of Δh is an improvement on current practice for estimating the rate of growth of productivity. In the remainder of this sub-section we explore the options available for estimating Δh .

3.1.1 Post treatment mortality

The only health measure routinely available from Hospital Episode Statistics is whether the patient is discharged dead or alive. For individual patients, death may or may not be directly due to the treatment that they received but aggregation over patients may convey some useful about the quality of treatment and how it is changing over time. If the function of hospital-based care is considered (in part) to be to save/prolong life, then at a system level it may be appropriate to consider mortality-based data as being broadly indicative of the extent to which this objective is being achieved.

Given the age of any patient their life expectancy can be determined from published actuarial life tables. It is thus possible to compute the number of life years lost by those who die in hospital.

Their health loss from mortality can be measured in one of two ways. The first approach simply assumes that their health status in all the future years s would have

been constant at $h(s) = 1$ and assigns $h(s) = 0$ to death. The second approach allows for the fact that an individual's health status varies as they age. Discounted quality adjusted life expectancy for an individual of a particular age *a* treated at date *t* is calculated as

$$(3) \qquad v(a,t) = \sum_{s=t}^{s=t+L(a,t)} \delta^{s-t} h(a,s)$$

by multiplying the quality adjusted health status index of a person aged *a* in year *s* by the appropriate discount factor and summing over`their life expectancy $L(a,t)$. (When health status varies with age it is better to calculate discounted QALYs achieved for death at different future ages and to use estimates of the conditional probabilities of death at varies ages to calculate the expected discounted quality. We will consider this refinement in phase 3.)

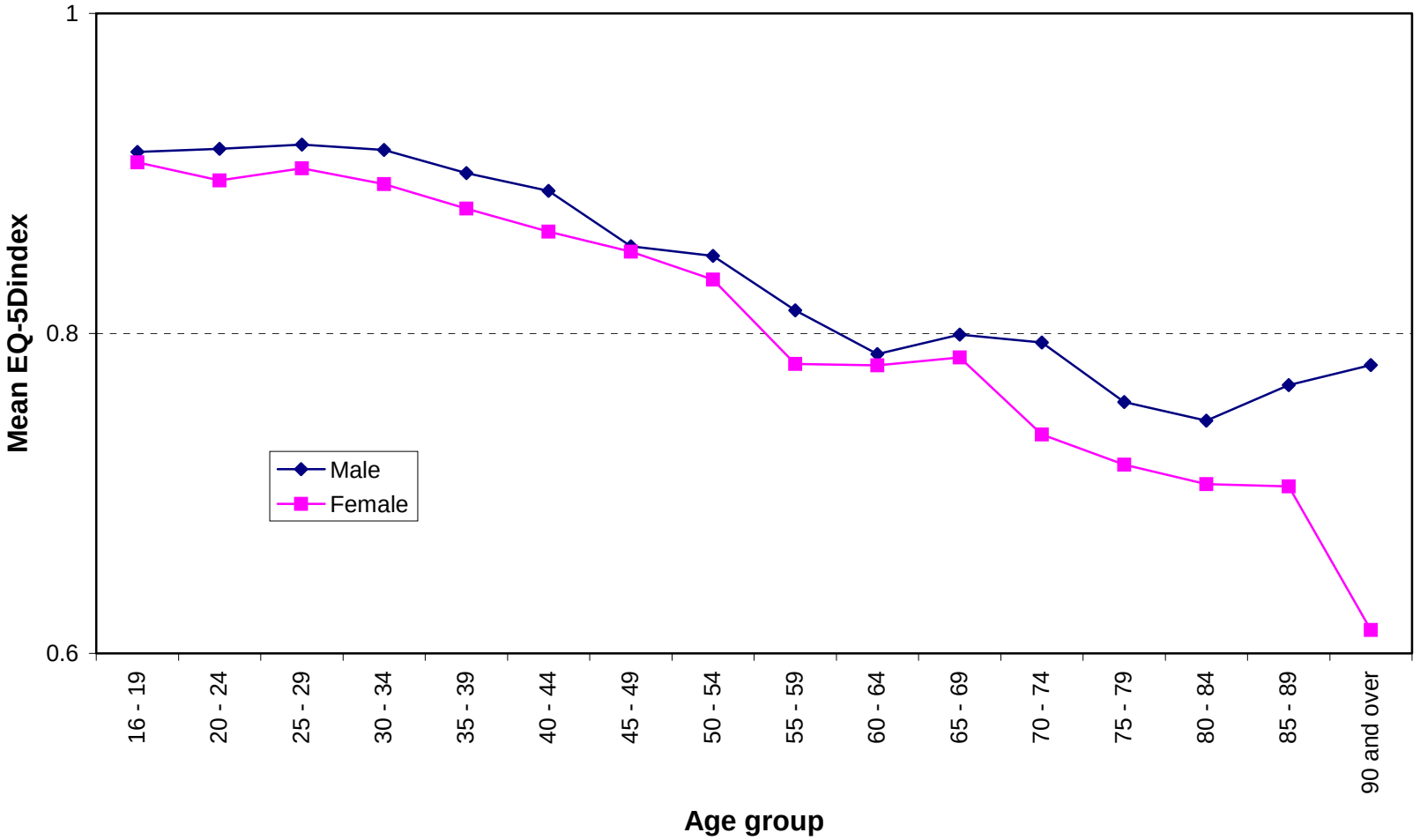
For each patient of given age and gender we can use period life expectancy estimated for males and females in the general population.[1] We derive the health status measures by using data from cross-section surveys of the population.

We use quality-adjustment data derived from applications of the EQ-5D generic health status instrument which measures health along five dimensions. The instrument is reproduced in Appendix 4. An overall health status index $EQ\text{-}5D_{vas}$ can derived using a visual analogue scale on a 0-100 scale where the endpoints are labelled "worst imaginable health" and "best imaginable health". Alternatively the index can be derived by time trade-off methods. A set of weights for the 5 EQ-5D dimensions can then be derived by analysis of the scores on the five dimensions describing a given health state and the overall summary scores for that health state for each individual. The resulting set of average weights can then be applied construct a health index $EQ\text{-}5D_{index}$ for all possible health states defined in EQ-5D. Such weights have been periodically determined in UK population studies. The set of weights most widely used in economic evaluation of healthcare are those from the Measurement and Valuation of Health Project (MVH) (Dolan *et al.,* 1994). To all intents and purposes these are the default values adopted in evaluation studies commissioned by NICE.

[1] Actuarial life tables published by ONS.

The 1996 Health Survey for England included EQ-5D as part of its interview schedule and the resulting EQ-5D$_{index}$ values for age and gender groups are shown in Figure 3.2.

Figure 3.2 Mean EQ-5D index values for the general population



The EQ-5D values can be combined with life expectancy data to generate a Quality-Adjusted Life Expectancy curve. If QALE is substituted for life expectancy then life years "lost" are expressed in terms of quality-adjusted life years "lost", i.e. QALYs.

We illustrate this procedure with a very simple worked example based on some preliminary analysis which will be refined and extended in phase 3. We use Hospital Episode Statistics data for the years 1996 and 2000. Patient-level episodes are categorised according to age and gender. 5-year age-groups are selected for patients older than 4 years on admission. For each age-group/gender cell the number of patient discharged dead was computed. These totals are given in Tables 3.1 and 3.2 for male and female admissions respectively. For example, in 1996 188,454 male patients aged 25 to 29 were admitted of whom 444 died. The corresponding age-adjusted mortality rate is thus 0.24%. Figures 3.3 and 3.4 plot the age-specific mortality rates for male and female admissions in 1996 and 2000. It will be noted that mortality rates in 2000 are consistently lower than the corresponding rates for 1996. This difference is reflected too in fall in the crude mortality rate for all admissions 2.66% to 2.28% (male) and 2.15% to 1.89% (female).

By taking the age-specific mortality rate for 1996 and applying it to the number of admissions in 2000 for the corresponding age groups, it is possible to estimate the number of patients expected to die among that year's admissions if they had experienced the 1996 mortality rate. The difference between this expected number of deaths and the observed number is presented in the final column of Tables 3.1 and 3.2. Compared to 1996, some 58,000 lives were "saved" in 2000 made up of 29,238 (male) and 28,826 (female).

Life expectancy for a male aged 27 (the midpoint of the 25-29 age band) was 48.95 years according to the ONS Actuarial Life Tables for 1996-1998. The total life years associated with the deaths recorded for each age group is given as the product of the period life expectancy and the number of patients discharged dead.. For male patients the total life years "lost" fell from 1,470,479 to 1,431,119 between 1996 and 2000. The corresponding totals for female patients are 1,542,240 and 1,492,216 life years.

In aggregate these mortality-based indicators represent a reduction of 8,025 in the

number of patients discharged dead, corresponding to 3.1% of the total recorded in 1996. When combined with life expectancy data the difference amounts to a reduction of 89,384 life years "lost", this being equivalent to 2.97% of the 1996 total.

Figure 3.3 Mortality rates for inpatients (male)

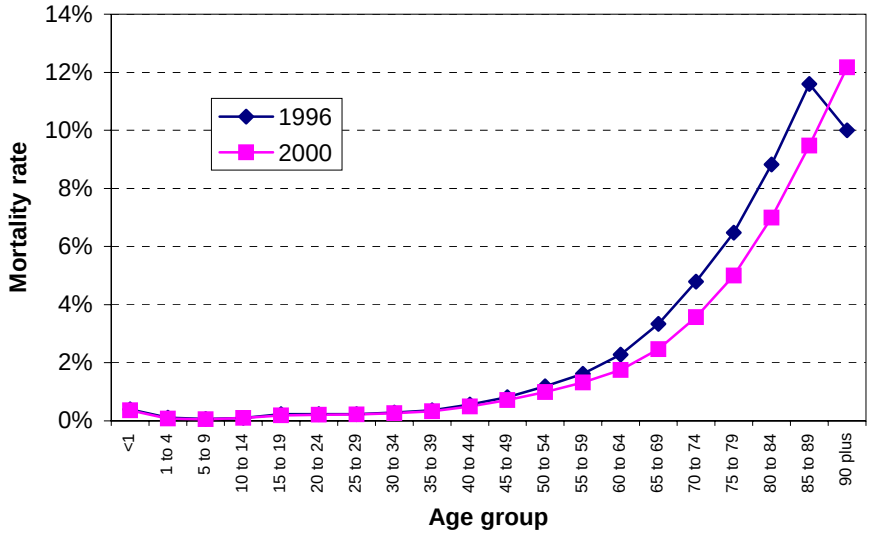


Figure 3.4 Mortality rates for inpatients (female)

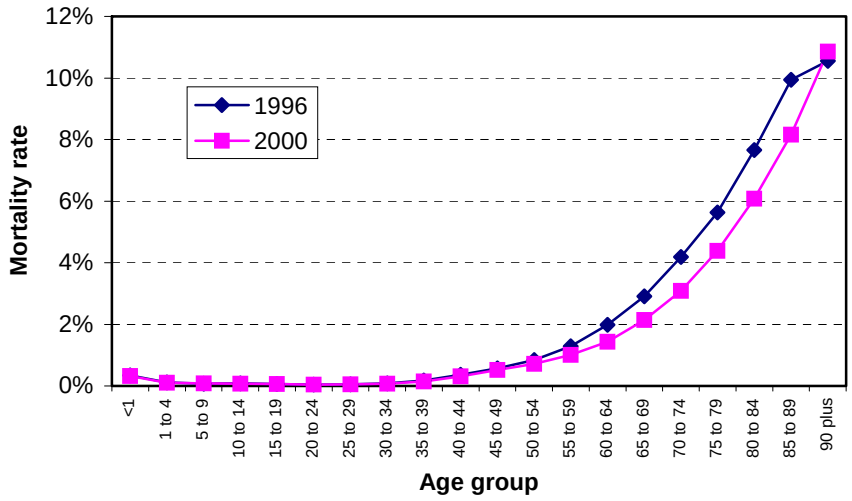


Table 3.1 Change in male hospital mortality 1996 - 2000

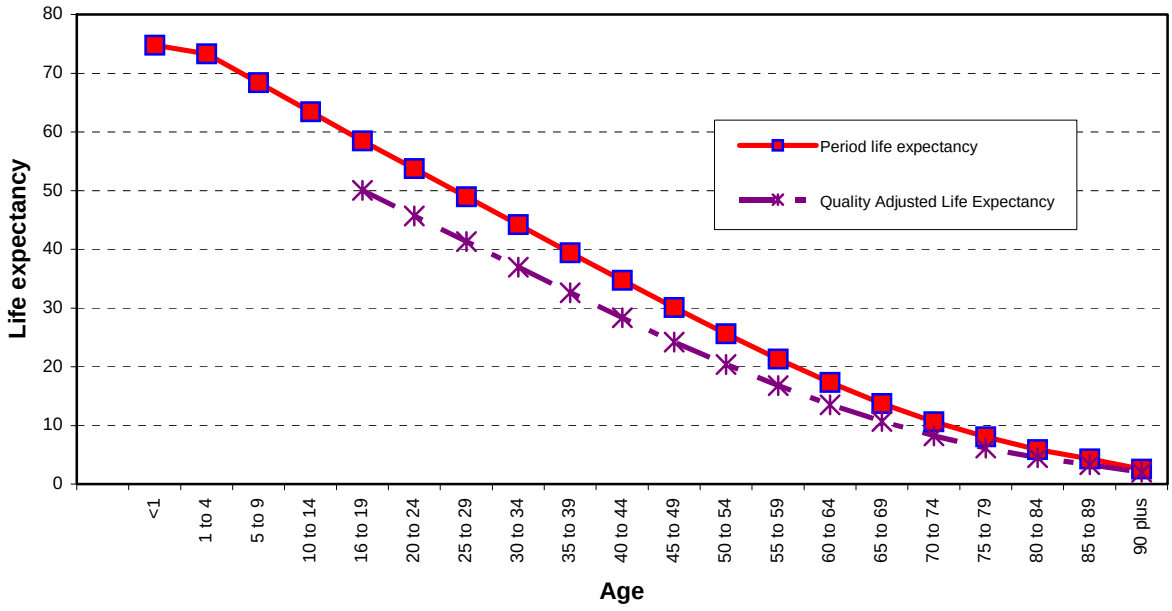
1996 Male						*2000 Male*						
Age	Total patient admissions	Discharged dead	Mortality rate %	Period life expectancy	Life years lost	**Age**	Total patient admissions	Discharged dead	Mortality rate %	Period life expectancy	Life years lost	2000 lives "saved"
<1	418,843	1,671	0.40%	74.76	124,924	**<1**	408,875	1,469	0.36%	75.62	111,086	162
1 to 4	207,423	227	0.11%	73.29	16,637	**1 to 4**	199,872	148	0.07%	74.12	10,970	71
5 to 9	160,066	111	0.07%	68.37	7,589	**5 to 9**	153,272	92	0.06%	69.18	6,365	14
10 to 14	122,520	123	0.10%	63.41	7,799	**10 to 14**	131,624	127	0.10%	64.22	8,156	5
15 to 19	121,240	274	0.23%	58.50	16,029	**15 to 19**	129,464	241	0.19%	59.30	14,291	52
20 to 24	148,620	345	0.23%	53.72	18,533	**20 to 24**	142,366	296	0.21%	54.50	16,132	34
25 to 29	188,454	444	0.24%	48.95	21,734	**25 to 29**	171,785	383	0.22%	49.72	19,043	22
30 to 34	211,093	613	0.29%	44.18	27,082	**30 to 34**	215,002	570	0.27%	44.96	25,627	54
35 to 39	201,362	741	0.37%	39.41	29,203	**35 to 39**	240,267	795	0.33%	40.20	31,959	89
40 to 44	194,491	1,092	0.56%	34.70	37,892	**40 to 44**	233,996	1,152	0.49%	35.49	40,884	162
45 to 49	233,348	1,893	0.81%	30.08	56,941	**45 to 49**	249,029	1,770	0.71%	30.86	54,622	250
50 to 54	252,618	3,012	1.19%	25.59	77,077	**50 to 54**	323,791	3,197	0.99%	26.38	84,337	664
55 to 59	283,502	4,601	1.62%	21.33	98,139	**55 to 59**	346,405	4,562	1.32%	22.08	100,729	1,060
60 to 64	338,052	7,716	2.28%	17.35	133,873	**60 to 64**	406,319	7,117	1.75%	18.06	128,533	2,157
65 to 69	388,506	12,961	3.34%	13.75	178,214	**65 to 69**	454,678	11,219	2.47%	14.36	161,105	3,950
70 to 74	414,721	19,863	4.79%	10.64	211,342	**70 to 74**	483,869	17,282	3.57%	11.11	192,003	5,893
75 to 79	352,080	22,792	6.47%	8.05	183,476	**75 to 79**	465,628	23,287	5.00%	8.39	195,378	6,856
80 to 84	260,056	22,932	8.82%	5.91	135,528	**80 to 84**	303,481	21,233	7.00%	6.14	130,371	5,528
85 to 89	137,519	15,949	11.60%	4.28	68,262	**85 to 89**	179,855	17,037	9.47%	4.44	75,644	3,822
90 plus	78,917	7,892	10.00%	2.56	20,204	**90 plus**	74,066	9,013	12.17%	2.65	23,884	-1,606
	4,713,431	**125,252**	**2.66%**		**1,470,479**		**5,313,644**	**120,990**	**2.28%**		**1,431,119**	**29,238**

Table 3.2 Change in hospital mortality: females 1996 - 2000

1996 Female						*2000 Female*						
Age	Total patient admissions	Discharged dead	Mortality rate %	Period life expectancy	Life years lost	**Age**	Total patient admissions	Discharged dead	Mortality rate %	Period life expectancy	Life years lost	2000 lives "saved"
<1	367,379	1,267	0.34%	79.76	101,056	**<1**	362,312	1,155	0.32%	80.34	92,793	95
1 to 4	140,864	160	0.11%	78.21	12,514	**1 to 4**	140,888	142	0.10%	78.78	11,187	18
5 to 9	115,730	89	0.08%	73.27	6,521	**5 to 9**	111,988	93	0.08%	73.84	6,867	-7
10 to 14	102,707	84	0.08%	68.30	5,737	**10 to 14**	108,381	83	0.08%	68.87	5,716	6
15 to 19	248,691	158	0.06%	63.36	10,011	**15 to 19**	263,924	163	0.06%	63.92	10,419	5
20 to 24	443,265	176	0.04%	58.46	10,289	**20 to 24**	426,805	183	0.04%	59.01	10,799	-14
25 to 29	626,344	331	0.05%	53.55	17,725	**25 to 29**	563,523	286	0.05%	54.10	15,473	12
30 to 34	590,641	468	0.08%	48.65	22,768	**30 to 34**	605,974	471	0.08%	49.20	23,173	9
35 to 39	381,285	674	0.18%	43.79	29,514	**35 to 39**	455,754	666	0.15%	44.34	29,530	140
40 to 44	265,907	973	0.37%	38.99	37,937	**40 to 44**	314,995	986	0.31%	39.53	38,977	167
45 to 49	284,630	1,619	0.57%	34.27	55,483	**45 to 49**	292,921	1,523	0.52%	34.81	53,016	143
50 to 54	279,229	2,341	0.84%	29.66	69,434	**50 to 54**	359,890	2,569	0.71%	30.20	77,584	448
55 to 59	261,718	3,379	1.29%	25.20	85,151	**55 to 59**	335,638	3,385	1.01%	25.72	87,062	948
60 to 64	275,479	5,475	1.99%	20.94	114,647	**60 to 64**	348,746	5,002	1.43%	21.44	107,243	1,929
65 to 69	321,131	9,326	2.90%	16.93	157,889	**65 to 69**	375,563	8,046	2.14%	17.37	139,759	2,861
70 to 74	369,895	15,486	4.19%	13.32	206,274	**70 to 74**	431,165	13,324	3.09%	13.65	181,873	4,727
75 to 79	368,619	20,743	5.63%	10.14	210,334	**75 to 79**	480,624	21,060	4.38%	10.38	218,603	5,986
80 to 84	356,344	27,300	7.66%	7.40	202,020	**80 to 84**	396,119	24,060	6.07%	7.56	181,894	6,287
85 to 89	255,991	25,450	9.94%	5.24	133,358	**85 to 89**	315,561	25,724	8.15%	5.33	137,109	5,648
90 plus	172,039	18,162	10.56%	2.95	53,578	**90 plus**	193,195	20,977	10.86%	3.01	63,141	-582
	6,227,888	**133,661**	**2.15%**		**1,542,240**		**6,883,966**	**129,898**	**1.89%**		**1,492,216**	**28,826**

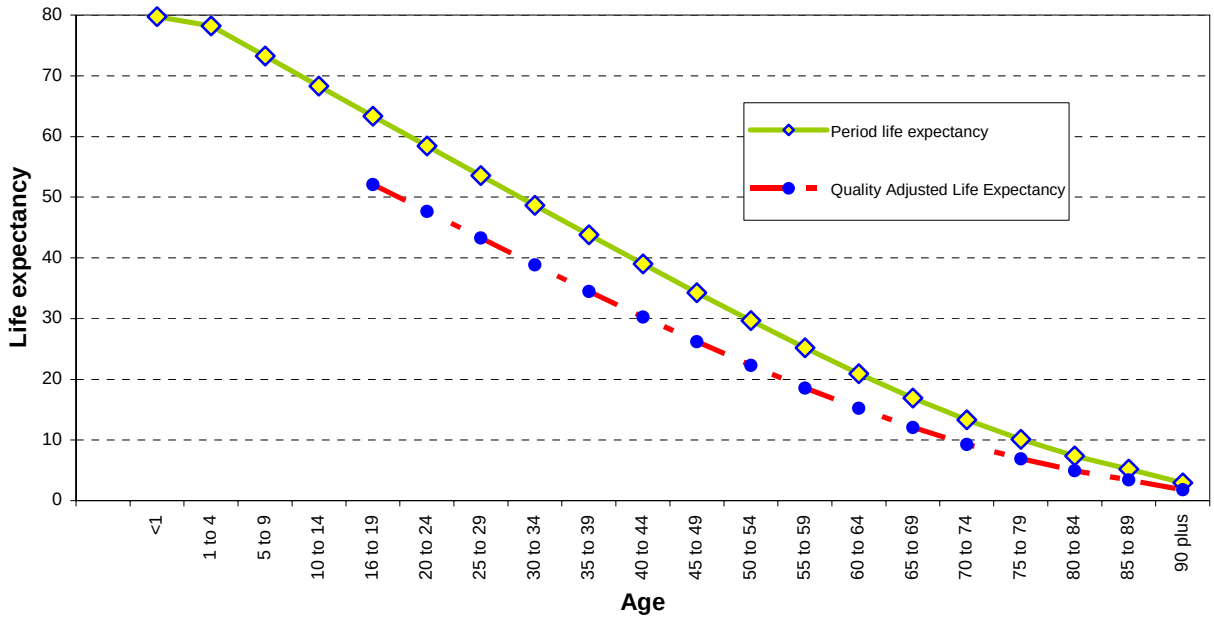
Figures 3.5a and 3.5b exemplify the results of combining period life expectancy data with mean EQ-5D$_{index}$ values obtained from a population survey conducted for the same year[2]. (We have not discounted future QALYs in this example.) In their unadjusted form, each 5-year period is associated with a mean life expectancy. For the 15-19 year old male this is 58.5 years. The upper line in each Figure represents the unadjusted period life expectancy. The lower line in each case represents the quality-adjusted life expectancy and is only plotted for adults aged 15 and over since general population EQ-5D scores were only available for this age range in the selected population survey. Data for younger respondents (as young as 8 years old) are available from other surveys.

Figure 3.5a Quality adjusted life expectancy (males: 1996)



[2] Health Survey for England : 1996

Figure 3.5b Quality adjusted life expectancy (females: 1996)



We can use this type of data to approximate the time profile of health status with treatment and hence to approximate $v^*(t)$. Thus for an individual of age a treated at date t we can calculate an expected discounted sum of QALYs as

(4) $$Ev^*(a,t) = [1-m(a,t)]v(a,t) + m(a,t)0 = [1-m(a,t)]\sum_{s=t}^{t+L(a,t)} \delta^{s-t} h(a,s)$$

where $m(a,t)$ is the mortality rate for the treatment of a person of age a at time t, $L(a,t)$ is life expectancy for a person aged a at time t, and we have assigned dead a health status value of zero. Summing over all ages gives an estimate of

(5) $$\sum_a Ev^*(a,t) = v^*(t)$$

The procedure has clear limitations:

- it does not provide any estimate of the without treatment health $v^o(t)$
- it uses values of the health index $v(a,t)$ from the general population, not the population of treated patients
- it uses hospital mortality rates $m(a,t)$ rather than treatment specific rates

- it uses general population life expectancy rather than treated population life expectancy
- it assumes a certain length of life rather than using age specific mortality probabilities to calculate expected discounted QALYs.

Nevertheless we believe that the approach, especially the derived estimates of the rate of change of $v^*(t)$, is an improvement on current methods which ignore health outcomes entirely. Even when surveys of patients provide before and after data on health status it will still be necessary to take account of patient mortality and the resulting lost QALYs by using mortality data.

In stage 3 we will extend and refine the analysis of mortality and EQ-5D data. We will

- use a series of cross section EQ-5D health surveys to allow for changing valuations of health states and will investigate the possibility of cohort effects
- extend the analysis to more years of HES data
- investigate whether results are sensitive to the use of mortality within 30 days of admission rather than mortality on discharge
- consider the use of uncertain length of life via age specific mortality rates rather than life expectancy.
- disaggregate to HRG level
- investigate whether it is possible to improve the estimates of life expectancy (and age specific mortality post treatment) by calculating treatment specific life expectancy and treatment and age specific mortality by linking HES and ONS death data

3.1.2 Clinical trial data

The use of published studies to estimate change in health status has a number of precedents (Berndt *et al*., 2001; Mai, 2004) and may be combined with or validated by expert opinion (Berndt *et al.*, 2002). In principle it can yield estimates of health status before treatment (h^b), and for those treated (h^*) and those not treated (h^o) at the same time point post treatment.

Nevertheless, there are some drawbacks to using data from published studies in constructing a productivity index:

- The study population may not be representative of the patients who receive the intervention in routine practice, given the exclusion criteria for many studies.
- The description of interventions in such studies may not map well to the classification of activities in the index. Usually trial interventions are very precisely defined, while fairly aggregated classifications are used in the construction of productivity indices.
- In most trials alternative interventions are compared, say of a new technology with existing practice. As such, there will be alternative estimates of effects. There are no clear grounds for favouring one estimate over the other in constructing a productivity index.
- The follow-up time in studies is variable, and hence the full treatment effect may not be captured by the estimate.
- Trials are not conducted for all NHS activities and, as such, only partial coverage can be achieved.
- Measurement of productivity change requires information about whether the effect of interventions has changed over time. There are very few examples of published studies that have attempted to ascertain the extent to which the effectiveness of the interventions being compared is determined by the time the intervention was made.

These caveats notwithstanding, we have sought to determine whether published studies contain information in the requisite form for use in constructing a productivity index. To do this we have surveyed users of the EQ-5D instrument and reviewed studies that have employed the instrument.

3.1.3 EQ-5D in clinical studies

The Outcomes Research Group in CHE provides central support for clinical users of EQ-5D in the UK, both inside the NHS and within the pharmaceutical industry. All contacts with the Group from users or potential users of EQ-5D in the UK and Republic of Ireland were contacted in order to establish how the instrument is

currently being used. Contacts were sent a short questionnaire requesting brief details of any studies they had conducted that included EQ-5D, together with details of any publications that resulted from those studies. Questionnaires were posted to 1,260 individuals. Replies were received from 319 individuals and a further 153 envelopes were returned having failed to be delivered. Of those who replied, 135 individuals provided details of their use of EQ-5D. A further request was sent to those who had replied positively to the initial questionnaire, asking for details of journal articles or other study reports. Replies were received from 62/135. Several respondents indicated that work was still in progress but many others referred to documentation that did not in fact report any EQ-5D data. This process resulted in 13 published papers being identified that reported EQ-5D.

Two other sources of data were examined to increase the coverage of reported usage of EQ-5D First, a literature search was conducted using Medline with the search terms *EuroQoL* and *EQ-5D.* This identified 464 papers of which 173 were published by researchers working in the UK. Of these, priority was given to those that were readily available in-house or via electronic download and to papers that reported on longitudinal studies with baseline and follow-up EQ-5D scores. Second, the Proceedings of the EuroQoL Group[3] were reviewed for similar reports.

In total 30 studies were identified through the three separate sources. The studies were reviewed, with information being extracted and summarised in a standard format. Full details for each of the studies is provided in Appendix 5 and the salient EQ-5D data are given in Table 3.3.

The main conclusions from this review are that:

1. EQ-5D has already been applied to the evaluation of a wide range of interventions suggesting that, in time, good coverage of NHS activity might be feasible.
2. There is a high degree of variability in the summary statistics reported in

[3] The EuroQoL Group convenes an annual Scientific Plenary meeting at which clinical applications and other methodological and developmental research findings are discussed. The most recent of these meetings was the 21st Scientific Plenary held in September 2004 at the Chicago Business School

studies, with differences in what measure of the average effect is used (mean or median) and whether or not variance is reported.

3. There is no common follow-up time across studies.
4. No studies report results according to the year in which the intervention took place so that even if there was more than one study of a particular treatment it would be difficult to use the to derive estimates of productivity growth.

We therefore will not be pursing this approach in stage 3 of the project, though we may use results from some of the studies to complement the more detailed analysis of some treatments which we will carry out with observational data.

Table 3.3 EQ-5D data from clinical study reports

Study number	Diagnostic Group	Treatment	Baseline EQ-5D score * (Std Deviation)	1st Follow-up score
1	Rheumatoid arthritis	Infliximab	0.43 (0.30) N=60	0.63 (sd not reported)
2	Ischaemic stroke	Stroke	0.31 (0.38) N=98	0.62 (0.33) N=98
3	Study A non psychotic psychiatric disorder; Study B rheumatoid arthritis	Study A brief psychodynamic interpersonal therapy/treatment as usual; Study B treatment as usual	Study A 0.36 (0.32) Study B 0.38 (0.37)	Study A 0.41 (0.32) Study B 0.43 (0.36)
4	Liver disease	Liver transplantation	0.53 (sd ?) N=152	0.59 (sd ?) N=152
5	Metastatic disease confined to liver		Resection 0.8 (sd ?) local ablation 0.9 (sd ?) irresectable 0.7 (sd ?)	Respectable 0.7 (sd ?) local ablation 0.7 (sd ?) irresectable 0.5 (sd ?)
6	Various. Largest categories were cardiovascular and respiratory	Referral to hospital at home or inpatient care	Not reported	Hospital at home 0.59 (0.15-0.78) Hospital 0.56 (0.19-0.73)
7	Cardiac surgery	Patients who had been allocated either to CABG-CPB or OPCAB surgery	$\underline{\text{EQ-5D}_{index}}$ CABG-CPB 0.82 (sd 0.25), n = 151 OPCAB 0.81 (sd 0.24), n = 161 $\underline{\text{EQ-5D}_{vas}}$ CABG-CPB 77.0 (sd 19.1), n = 151 OPCAB 76.0 (sd 16.1), n = 165	
8	Chest pain	Chest pain observation unit or routine care	$\underline{\text{EQ-5D}_{vas}}$ Obs unit = 79.8 Routine care = 75.7	$\underline{\text{EQ-5D}_{vas}}$ Obs unit = 77.9 Routine care = 71.86
9	Chest pain	Chest pain observation unit	EQ-5D_{index} = 0.63 n = 166	EQ-5D_{index} = 0.69 n = 110
10	Lower leg injury	Fasciotomy within 12 hours of injury	Impact on function EQ-5D_{vas} =71.99 EQ-5D_{index} = 0.619 No impact on function $\text{EQ-5D}_{vas\_}$ = 95.1 $\text{EQ-5D}_{index\_}$ =0.946 Skin graft EQ-5D_{vas} 74.1 EQ-5D_{index} =0.619 No skin graft EQ-5D_{vas} = 89.0 EQ-5D_{index} = 0.946 Appearance some problem EQ-5D_{vas} = 74.2 EQ-5D_{index} =0.631 Appearance not a problem EQ-5D_{vas} = 93.5 EQ-5D_{index} = 0.931	
11	Lower leg injury	Assessment following surgery. Mean follow-up time of final	EQ-5D_{vas} = 68	

		review was 46 months (15-80)		
12	Arthritis	Intervention Group received Arthritis Self-Management Programme i.e. six weekly sessions each lasting approx 2h, delivered by pairs of lay leaders trained by Arthritis Care	Intervention Group EQ-5D$_{index}$ = 0.43 EQ-5D$_{vas}$ = 56.78 Control Group EQ-5D$_{index}$ = 0.44 EQ-5D$_{vas}$ = 57.87	Intervention group EQ-5D$_{index}$ = 0.47 EQ-5D$_{vas}$ = 57.12 Control group EQ-5D$_{index}$ = 0.45 EQ-5D$_{vas}$ = 59.04 Difference n/s
13	Spinal pain	Usual care or usual care + referral to osteopathy clinic for three or four sessions of spinal manipulation and advice	Intervention Group EQ-5D$_{index}$ = 0.56 EQ-5D$_{vas}$ =58.9 Control Group EQ-5D$_{index}$ = 0.50 EQ-5D$_{vas}$ = 61.3	Intervention group EQ-5D$_{index}$ = 0.67 EQ-5D$_{vas}$ = 69.5 Control group EQ-5D$_{index}$ = 0.56 EQ-5D$_{vas}$ = 66.1
14	Carers of older patients with mental health problems	n/a	EQ-5D$_{index}$ = 0.70 EQ-5D$_{vas}$ =67.36	
15	People discharged from hospital after acute myocardial infarction	none	EQ-5D$_{index}$ = 0.68	not reported
16	Aged over 79 years	Home-based medication review by pharmacists	EQ-5D$_{index}$ =0.61	not reported
17	Abstaining alcoholics	Inpatients at a voluntary sector residential alcohol treatment centre, abstaining from alcohol	EQ-5D$_{vas}$ = 54.1	EQ-5D$_{vas}$ = 67.5
18	Porphyria	n/a	Mean EQ-5D$_{index}$ ranged from 0.10 for 70-79 year old males to 0.86 for 20-29 year old females. Mean EQ-5D$_{vas}$ ranged from 35.0 for 70-79 year old males to 77.5 for 30-39 year old males. Patients with manifest symptomology had EQ-5D$_{index}$ scores of 0.71 compared with those with latent symptomology of 0.76.	
19	Elderly people with balance problems	Enhanced Balance Training (EBT) or conventional physiotherapy (CT)	EBT group mean EQ-5Dvas = 57.8 +/- 19.7 CT group mean EQ-5D$_{vas}$ = 59.4 +/ 17.2	EBT group mean VAS 65.1 +/- 19.6 CT group mean VAS 64.9 +/ 17.3
20	Women with urodynamic stress incontinence	Either colpsuspension or tension-free vaginal tape	T-f vag tape EQ-5D$_{index}$ = 0.78 [0.71 - 0.92] Colposusp. EQ-5D$_{index}$ = 0.79 [0.71 - 0.92]	T-f vag tape 0.79 [0.71 - 0.92] Colposusp. 0.75 [0.69 - 0.88]
21	Elderly patients with an acute illness requiring hospital admission	One of two settings, community or district general hospital	DGH group Mean EQ-5D$_{vas}$ = 53.9 (50.7-57.1) Median EQ-5D$_{index}$ =0.36 (0.07-0.69) CH group EQ-5D$_{vas}$ = 50.5 (47.5-53.6 Median EQ-5D$_{index}$ = 0.26 (0.05-0.69)	
22	Parkinsons' disease	Either nurse specialist care or standard care	Mean EQ-5D$_{index}$ = 0.47 (0.35)	Mean INDEX Nurse Group 0.37 |(0.35)

		(control group)		Control Group 0.39 (0.35)
23	n/a	Influenza vaccination or placebo	low HADS group EQ-5D$_{vas}$ = 84..5 (14.4) high HADS group EQ-5D$_{vas}$ = 63.1 (17.4) Difference = sign. (p<0.001)	Mean (SD) VAS for low HADS group 83.2 (12.7) for high HADS group 58.9 (18.4) Difference = sign. (p<0.001)
24	Low back pain	Intervention group received usual care and radiograph of the lumbar spine at their local hospital. Control group received the usual care from their doctor, including radiography if the doctor considered it to be clinically necessary	Intervention Group Median EQ-5D$_{index}$ = 0.69 (interquartile range 0.62 - 0.76) Median EQ-5D$_{vas}$ =70 (50-80) Control group Median EQ-5D$_{index}$ = 0.69 (0.62-0.76) Median EQ-5D$_{vas}$ = 70 (50-80)	Median EQ-5D INDEX (interquartile range) Intervention Group = 0.80 (0.69 - 0.88) Control group = 0.80 (0.69-0.91) Median VAS (interquartile range) Intervention group 75 (60-90) Control group 80 (70-90)
25	Gynaecology	Laparoscopic hysterectomy or standard hysterectomy (either abdominal or conventional according which was most appropriate) giving two trials vaginal versus laparoscopic and abdominal versus laparoscopic hysterectomy 2;1 randomisation in favour of the laparoscopy	Vaginal Trial Mean EQ-5D$_{index}$ Lap group 0.75, vag group 0.76. Abdominal trial Lap group 0.72 abdom group 0.69	Mean EQ-5D INDEX: Vaginal Trial, Lap group 0.88, vag group 0.85. Abdominal trial Lap group 0.83 abdom group 0.83
26	Sleep disorders	Nasal continuous positive airways pressure	Mean EQ-5D$_{index}$ = 0.79 (0.21)	Mean EQ-5D$_{index}$ = 0.84 (0.25)
27	Sleep disorders	Continuous positive airways pressure therapy	Mean EQ-5D$_{index}$ = 0.78(0.22) Mean EQ-5D$_{vas}$= 66.57 (18.91)	Mean EQ-5D$_{index}$ = 0.83(0.22) Mean EQ-5D$_{vas}$= 71.72 (18.12)
28	Benign prostate hypertrophy	Transurethral resection of the prostate (TURP) or lasor vaporization prostatectomy	TURP group Mean EQ-5D$_{index}$ = 0.81(0.18) Mean EQ-5D$_{vas}$ = 78.3 (13.2) Laser group Mean EQ-5D$_{index}$ = 0.81(0.18) Mean EQ-5D$_{vas}$ =75.8(17.1)	TURP group Mean EQ-5D$_{index}$ = 0.85(0.17) EQ-5D$_{vas}$ = 79.9 (16.3) Laser group Mean EQ-5D$_{index}$ =0.85(0.20) EQ-5D$_{vas}$ = 74.2(19.5)
29	Acute lower back pain	IFT painful area and 'The Black Book, IFT spinal nerve and 'The Black Book OR Control group 'The Black book' only	Painful area group Median EQ-5D$_{index}$ = 0.69, Spinal Nerve group Median EQ-5D$_{index}$ = 0.76, Control group Median EQ-5D$_{index}$ = 0.69	Painful area group Median EQ-5D$_{index}$ = 0.80 Spinal Nerve group Median EQ-5D$_{index}$ = 0.79 Control group Median EQ-5D$_{index}$ = 0.93
30	Varicose veins	PIN stripping (43 patients) or Conventional stripping (37)	PIN group Median EQ-5D$_{index}$ = 0.73 (0.66-0.83) Conventional group Median EQ-5D$_{index}$ = 0.8 (0.69-1.0)	PIN group Median EQ-5D$_{index}$ = 0.8 (0.73 - 1.0) Conventional group Median EQ-5D$_{index}$ = 0.83 (0.69-1.0)

EQ-5D$_{index}$ unless otherwise stated

3.1.4 Observational data

In view of the difficulties associated with relying on clinical trial data, we have recommended to the Department of Health pre and post treatment health status should be measured using surveys of the general stream of NHS patients (see Data Memorandum in Appendix 1). In the absence of such general patient population data we have identified three existing sources of more limited observational data:

- Health Outcomes Data Repository
- York District Hospital
- BUPA

These sources cannot yield information for a calculation of productivity growth in the secondary care sector as a whole but they can be used to examine the usefulness of such data, to guide the specification of more general surveys and to triangulate estimates of productivity growth obtained from other sources for some HRGs

3.1.4.1 Health Outcomes Data Repository

Since June 2002, the Health Outcomes Data Repository (HODaR) (www.cardiff-research-consortium.co.uk/hodar) has operated a continuous health status survey of all inpatients and outpatients at a single large Welsh Trust. These are linked to individual-level hospital, primary and community care data. We have been provided with the HODaR database, containing 29,541 observations. We have analysed the information in order to gain an understanding of the challenges faced in collecting and using observational data.

Data on health status are collected using two generic health survey instruments;

1. RAND MOS SF-36. The SF36 provides multiple measures of different aspects of health outcome but it is not a simple matter to form a single aggregate health status measure for use in an output index.
2. EQ-5D. The EQ-5D is designed to produce a single index and its five dimensions have been calibrated in terms of social preference weights of a UK population. The instrument is shown in Appendix 4. We report here some of the data in HODAR on EQ-5D scores.

Inpatients are sent a postal questionnaire six weeks post discharge. Outpatients are surveyed at the time of their clinic appointment.

Because the survey is linked to HES data, it is possible to analyse the health status survey by condition, defined by ICD-10 diagnoses codes, procedure codes or Healthcare Resource Group. Table 3.4 presents the mean and standard deviation in EQ-5D score by HRG for all 29,541 patients in descending order of HRG frequency (only HRGs containing more than 65 patient surveys are listed). For the full sample, mean EQ-5D score is 0.66, implying that the average health state experienced by these patients has 66% of the value of being in full health. The variation in EQ-5D score is high, with a standard deviation of 0.32. This implies caution in the use of the mean estimates as precise measures of the health states experienced by patients in each HRG.

Table 3.4 EQ-5D score by HRG

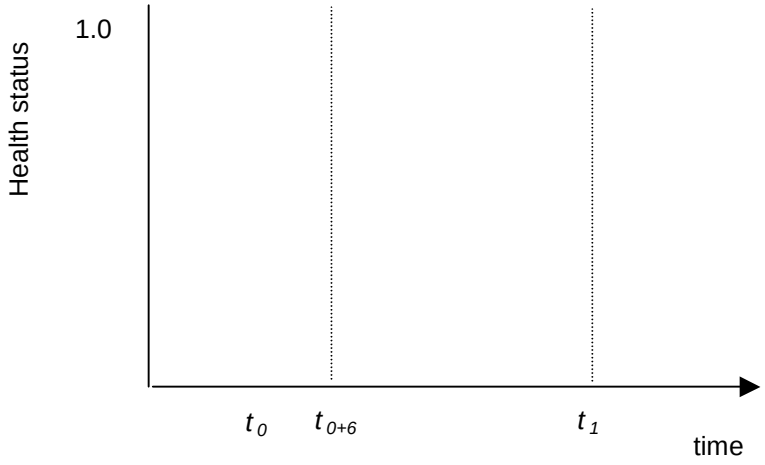
Group	N	Mean EQ-5D	Standard Deviation in EQ-5D	Group	N	Mean EQ-5D	Standard Deviation in EQ-5D
Total	29541	0.66	0.32	**E03**	120	0.69	0.29
E14	1141	0.61	0.31	**M09**	120	0.82	0.30
J37	811	0.76	0.28	**Q14**	120	0.56	0.33
F06	741	0.69	0.30	**F54**	118	0.77	0.26
B02	700	0.67	0.29	**H37**	116	0.63	0.27
M06	631	0.74	0.30	**L27**	115	0.66	0.32
S22	550	0.63	0.31	**L46**	114	0.59	0.34
F35	547	0.73	0.29	**C22**	112	0.72	0.33
M07	515	0.74	0.29	**F32**	112	0.73	0.26
L21	479	0.69	0.30	**B03**	110	0.60	0.31
N12	422	0.85	0.23	**M11**	105	0.85	0.24
J02	387	0.73	0.26	**S25**	105	0.57	0.41
E36	338	0.63	0.34	**F31**	103	0.70	0.27
E15	318	0.65	0.32	**S01**	99	0.70	0.26
F47	315	0.69	0.31	**C02**	93	0.77	0.29
F16	310	0.67	0.30	**Q12**	93	0.58	0.31
H10	296	0.59	0.33	**R02**	92	0.43	0.38
H04	253	0.51	0.30	**E10**	90	0.61	0.32
M03	243	0.70	0.28	**F82**	89	0.85	0.25
C24	237	0.70	0.31	**H06**	89	0.43	0.33
M02	232	0.75	0.28	**M01**	89	0.81	0.25
F74	228	0.77	0.26	**H19**	87	0.64	0.35
N07	225	0.89	0.20	**J03**	87	0.77	0.28
E04	221	0.63	0.27	**B06**	82	0.71	0.27
L17	217	0.74	0.27	**F37**	81	0.61	0.37
H02	204	0.53	0.30	**F93**	81	0.80	0.28
H17	199	0.67	0.32	**H40**	81	0.74	0.30
E12	196	0.68	0.28	**A22**	80	0.54	0.33
E33	194	0.58	0.28	**H26**	78	0.42	0.34
A07	191	0.25	0.34	**L26**	77	0.73	0.33
G14	190	0.72	0.29	**E09**	76	0.68	0.29
E34	178	0.57	0.32	**F56**	76	0.64	0.36
B09	172	0.69	0.30	**F42**	75	0.61	0.39
B05	164	0.66	0.31	**H33**	75	0.49	0.31
J35	154	0.77	0.28	**C01**	74	0.65	0.36
F73	151	0.70	0.26	**K16**	74	0.65	0.34
M05	149	0.79	0.28	**E31**	73	0.62	0.32
S16	147	0.55	0.38	**K01**	72	0.71	0.28
E30	142	0.71	0.29	**E18**	70	0.53	0.35
H13	142	0.66	0.32	**L10**	70	0.71	0.32
Q11	142	0.80	0.24	**Q10**	70	0.71	0.30
D21	140	0.56	0.34	**A30**	69	0.55	0.37
L19	140	0.70	0.28	**D13**	68	0.51	0.38
D20	136	0.45	0.34	**E32**	68	0.71	0.30
E35	133	0.57	0.31	**J04**	68	0.77	0.24
F46	128	0.60	0.32	**L53**	68	0.76	0.28
E29	126	0.67	0.24	**R03**	68	0.31	0.36
E08	125	0.64	0.28	**S98**	68	0.63	0.31
F36	123	0.52	0.34	**D02**	67	0.61	0.28
H14	123	0.76	0.28	**K04**	67	0.67	0.37
C14	121	0.76	0.32	**L43**	67	0.76	0.26

For 2,587 patients who had two courses of treatment, two EQ-5D surveys are available so that it is possible to measure the change in their health status over time Although such patients are unlikely to be typical of all patients receiving care, we have concentrated on analysing data for this subset of patients to see what light it sheds on changes in health status.

This subset of patients differ along three important dimensions:

1. The elapsed time between the two surveys. This can be calculated because we have information on the date that each survey was completed.
2. Whether the two surveys relate to the same underlying condition. We have assessed whether coding of diagnosis remains similar for the two surveys. Different diagnostic codes may indicate that the surveys relate to different underlying conditions.
3. The ordering of the settings to which the surveys refer. Ordering is determined by the survey identification number. The interpretation placed on the change in health status observed between each survey depends on this ordering. There are four possible scenarios.

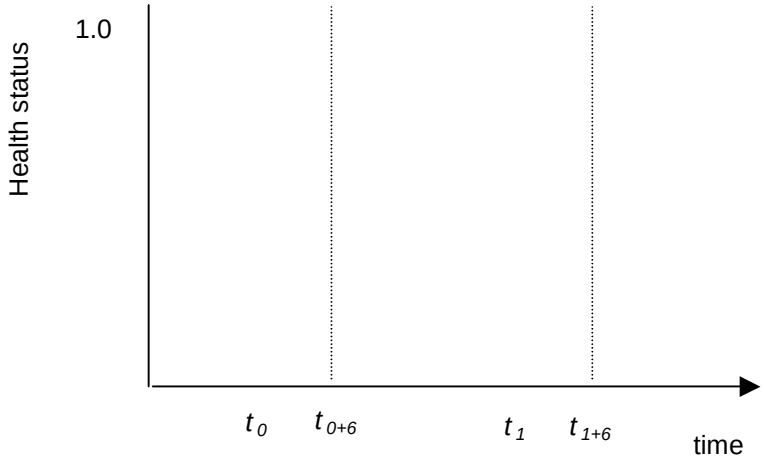
Scenario 1: Inpatient survey followed by outpatient survey.



Under this scenario, the patient is admitted at t_0, completes a post-discharge survey at t_{0+6} and completes an outpatient survey during their clinic visit at time t_1. The change in health status is calculated as $\Delta h_1 = EQ5D_{t_1} - EQ5D_{t_{0+6}}$. Even if the outpatient visit

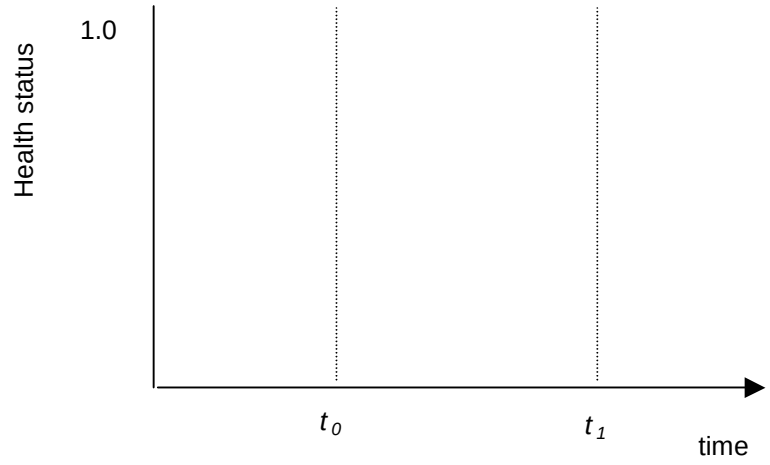
relates to the same health problem as the earlier admission it is not possible to predict the sign of Δh_1. It may be positive if the patient continues to recover during the time between the surveys. But if the outpatient visit is triggered by deterioration in the patient's condition, Δh_1 will be negative. Moreover the full treatment effect will not be captured, because any recovery occurring between t_{0+6} and t_0 is not recorded. It may well be that, for many interventions, the most dramatic changes in health status occur during this period.

Scenario 2: Two inpatient surveys



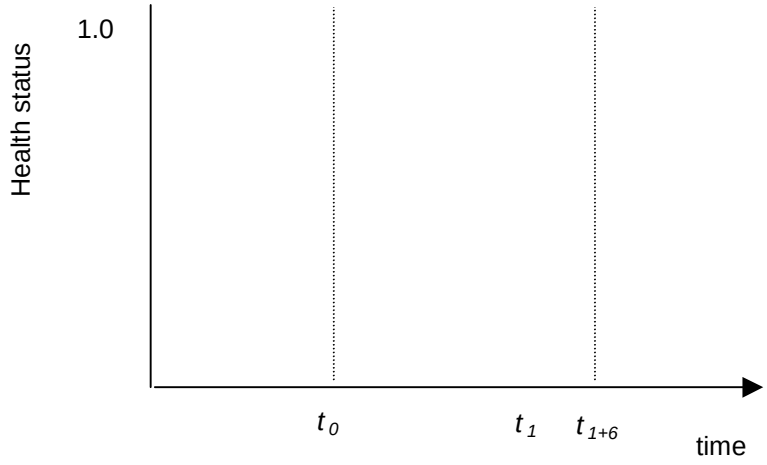
Under this scenario the change in health status is calculated as $\Delta h_2 = EQ5D_{t_{1+6}} - EQ5D_{t_{0+6}}$. There is a high probability that the two inpatient admissions are unrelated, this probability increasing the longer the elapsed time between the two surveys. Irrespective of whether or not the admissions are related, it is not possible to predict the sign of Δh_2.

Scenario 3: Two outpatient surveys



Here the change in health status is calculated as $\Delta h_3 = EQ5D_{t_1} - EQ5D_{t_0}$. As for the previous scenario it is not possible to predict the sign of Δh_3.

Scenario 4: Outpatient survey followed by inpatient survey



The change in health status is calculated as $\Delta h_4 = EQ5D_{t_{1+6}} - EQ5D_{t_0}$. This scenario may represent a before-and-after measurement of the intervention effect. This is more likely if the outpatient visit and admission are related, the shorter the lapsed time between the outpatient visit and admission, and if most of the post-treatment recovery occurs in the six-week post-discharge period. The more these conditions prevail, the more likely that Δh_4 will have a positive sign.

Table 3.5 shows the number of patients, change in EQ-5D score, and time between the

two surveys for each of the above scenarios. The mean change in EQ-5D is negative for all four scenarios, although there is a wide range in scores. There are a number of EQ-5D health states to which negative values are attached, implying that these are health states considered worse than death. Changes into or out of such health states may lead to Δh taking values in excess of ± 1.

Table 3.5 Change in EQ-5D score and time between surveys, by scenario

Scenario	N	Mean change in EQ-5D	Min change in EQ-5D	Max change in EQ-5D	Mean time between surveys (days)
1 Inpatient to Outpatient	502	-0.0075	-1	1	633
2 Inpatient to Inpatient	1612	-0.0015	-1	1.016	634
3 Outpatient to Outpatient	192	-0.0196	-0.912	0.796	489
4 Outpatient to Inpatient	281	-0.0069	-1	1	581

The lapsed time between surveys is often considerable, amounting to around a year and a half on average. This is the case even for scenario four, where the average time between surveys amounts to 581 days. This undermines the case for considering the surveys for this subset of patients as constituting before-and-after measurements. That said, the distribution over time for these patients is highly skewed. The patients can be classified into three groups:

1. 48 patients for whom the date of the outpatient and inpatient surveys are identical (the ordering of surveys is determined by the survey identification number). This implies that the outpatient survey was completed retrospectively, at the same time as the inpatient survey. There may be problems of recall for these patients.
2. Those for whom the elapsed time is very long. It is doubtful that these surveys constitute before-and-after measurements.
3. Remaining patients, where elapsed time is short enough to suggest they may be before-and-after surveys. Of course, this requires a judgement to be made about what constitutes a ‘short enough’ time.

The possibility that the two surveys constitute before-and-after measurement is further undermined when considering the diagnosis recorded at each survey. The ICD-10 codes recorded over the two periods are rarely the same, and there is little consistency

even in the HRG chapter to which the patient is classified. Table 3.6 shows how the 281 patients under scenario four are classified to HRG chapters at the first and second survey. Shaded boxes along the diagonal indicate the number of patients who remain in the same chapter for both surveys with only 43% doing so.

The off-diagonal cells show the number of patients who moved HRGs between the surveys. As an example, the first row shows where the 7 patients originally classified to Chapter A (Nervous System) were classified in their second survey. Thus, 3 remained in this chapter, one patient was subsequently classified to Chapter C (Mouth, Head, Neck and Ears) and three to Chapter H (Musculoskeletal). The extent of this movement across HRG chapters implies that caution should be exercised in attaching the change in EQ-5D score to a specific underlying condition. Moreover, inconsistency in the condition to which patients are classified over time increases data requirements substantially (by the power of the additional classes that need to be considered).

Table 3.6 Change in HRG chapter classifications between surveys, scenario 4 patients

	HRG Chapter (2nd Survey)																	
HRG Chapter (1st Survey)	A	B	C	D	E	F	G	H	J	K	L	M	N	Q	R	S	U	Total
Nervous System	3		1					3										7
Eyes and Periorbita		1			1	2					1	1					2	8
Mouth, Head, Neck and Ears	1		1			1								1		1		5
Respiratory System		1		6	2	1			1								2	13
Cardiac Surgery		1	1		34	2	1	3			4	1		2			4	53
Digestive System			2		2	18	4	4	3	2	3	3		2	1	3	5	52
Hepato-Biliary and Pancreatic System		1				1	5											7
Musculoskeletal System		1			4	2		9	1		2			2	1	3	2	27
Skin, Breast and Burns	1		1		3	3		1	4		1	1					2	17
Endocrine and Metabolic System					2			1	1	1							2	7
Urinary Tract and Male Reproductive System					2			1	1		11	1	1			2	2	21
Female Reproductive System		2	1		2	3			1			11		1		1	3	25
Obstretics and Neonatal Care						1						1	1		1			4
Vascular System			1		1									2			1	5
Spinal Surgery and Primal Surgery Conditions													1					1
Haematology, Infectious Diseases					1	1			1		1					5	1	10
Mental health					1												1	2
Invalid Primary Diagnosis					2				1		1	1			1	1	10	17
Total	5	7	8	6	57	35	10	22	14	3	24	20	3	10	4	16	37	281

Table 3.7 Change in EQ-5D score between surveys, scenario 4 patients

	HRG Chapter (2nd Survey)																
HRG Chapter (1st Survey)	A	B	C	D	E	F	G	H	J	K	L	M	N	Q	R	S	U	Total
Nervous System	0.20		-0.16					-0.39										-0.11
Eyes and Periorbita		-0.04			0.15	-0.06					-0.33	0.11					-0.16	-0.07
Mouth, Head, Neck and Ears	-0.10		0.59			-0.84								0.00		-0.18		-0.11
Respiratory System		-0.59		0.06	0.06	0.00			-0.10								0.31	0.03
Cardiac Surgery		-0.10	-0.03		0.04	0.12	0.04	-0.09			0.05	0.75		0.19			0.04	0.05
Digestive System			-0.30		-0.28	-0.03	-0.28	0.08	-0.06	-0.40	0.01	0.18		0.11	0.56	0.13	-0.07	-0.04
Hepato-Biliary and Pancreatic System		-0.16				-0.21	0.08											0.00
Musculoskeletal System		-0.07			-0.11	-0.14		0.19	0.00		0.11			0.06	-0.19	0.06	-0.02	0.05
Skin, Breast and Burns	-0.28		-0.10		0.03	0.12		0.15	-0.17		-0.07	-0.24					0.00	-0.05
Endocrine and Metabolic System					-0.35			-0.26	0.00	0.00							0.00	-0.14
Urinary Tract and Male Reproductive System					0.05			0.00	0.00		-0.12	0.00	1.00			0.00	-0.08	-0.02
Female Reproductive System		-0.04	-0.07		-0.03	0.08			-0.54			-0.12		0.00		1.00	0.04	-0.03
Obstretics and Neonatal Care						0.00						0.20	-0.15		0.35			0.10
Vascular System			-0.15		-0.31									0.15			0.00	-0.03
Spinal Surgery and Primal Surgery Conditions													0.00					0.00
Haematology, Infectious Diseases					-0.66	1.00			-0.07		-0.06					-0.25	0.00	-0.10
Mental health					0.10												0.00	0.05
Invalid Primary Diagnosis					0.35				0.00		0.07	0.20			-0.62	0.49	0.03	0.07
Total	0.04	-0.15	-0.07	0.06	0.00	-0.01	-0.07	0.02	-0.11	-0.27	-0.05	0.01	0.28	0.10	0.02	0.04	0.01	-0.01

Finally for completeness, Table 3.7 shows the mean change in EQ-5D score (Δh_4) for the patients in scenario 4 according to their classifications to HRG chapters across the two surveys. Obviously, small numbers in each cell preclude drawing any conclusions from these data.

We have analysed the HODaR set of observational data to ascertain whether the information can be utilised in the construction of outcome weights for a productivity index. We have concluded that the HODaR data are unsuitable for this purpose. The surveys have not been administered with the express intention of collecting before and after information. Although multiple surveys exist for a subset of patients, it is unlikely that many of these constitute before-and-after measurements. However, the HODaR data do provide an indication of the variation in EQ-5D scores for particular conditions (e.g. specific diagnoses or HRGs). This information might be used to assess the sample size requirements for the collection of other observational data.

3.1.4.2 York District Hospital

The Orthopaedics and Trauma unit at York District Hospital has been collecting SF36/12 data since March 2001. The patient health state is first recorded at the pre-op outpatient appointment. Post-operative outcomes are obtained from postal questionnaires and recorded at intervals of three, six, twelve and twenty-four months The data set is small, with 253 patients recorded at the pre-op stage. Numbers fall with the length of follow up: 203 patients after three months but only 25 returns after twenty-four months (noting that data collection commenced only three years ago, so two year follow-up is applicable only for a proportion of the sample). The DH is arranging for the Trust to continue collecting these data to enable us to examine it more fully in stage 3 of the project. For a few orthopaedic procedures the data should give us a better indication of pre and post treatment outcomes than existing clinical trial data. The information should also permit initial analysis of any trend changes in outcomes over time. Given the importance of increasing the collection of routine outcomes data by the NHS, the experience of York District Hospital can also be used to examine the most efficient ways of collecting routine outcomes data, though the particular instrument employed in York may not the most suitable for a national sample of NHS patients.

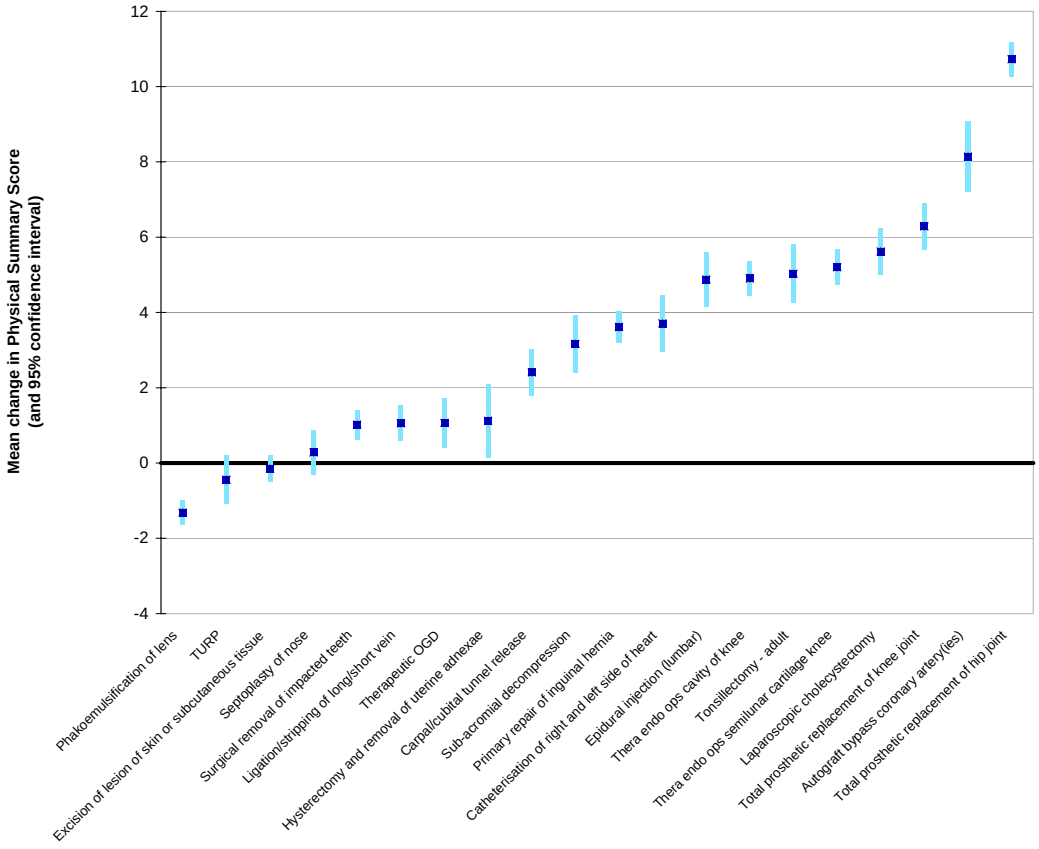
3.1.4.3 BUPA

As we noted in our First Interim Report, BUPA have been collecting outcomes data (before and after health status) on their patients. We have recently obtained a dataset from them which we intend to analyse in the next stage of the project. BUPA changed database supplier in January 2003 and we have been advised that this may have led to a change in data reliability.

BUPA have collected the data for the last 6 years from 70 private hospitals and it contains 90,000 patient treatment episodes. Patients fill in a health status questionnaire before treatment and three or four months post treatment depending on whether the instrument is SF-36 or VF-14 (for cataract procedures since 2001). Since 2002 BUPA has concentrated on collecting data from 20 sentinel episodes (high volume pivotal procedures), including primary inguinal hernia, CABG, PTCA, primary hip replacement and cataracts. BUPA have undertaken some analysis of these data. For example, Figure 3.6 reports the change in SF36's physical summary score for selected procedures.

The primary aim of the analysis will be to identify underlying general trends in changes in health status for particular procedures as a means of triangulating other estimates of quality improvements due to technological progress in treatment. The health gain for an individual undergoing a procedure or group of procedures would be regressed on a set of covariates and on time dummies. The coefficients on time dummies will identify trends in the health effects of treatment. Potential covariates include age, gender, and insurer dummy variables. There are also data on the patient area of residence (at three digit postcode level) which may be used to attribute socio-economic variables. In addition, we will include a set of anonymised hospital dummies to allow for changes in the mix of provider units over time.

Figure 3.6 Changes in BUPA patients' SF-36® physical summary score after 12 weeks, by surgical intervention



Source: Vallance-Owen (2004)

Although only preliminary investigation has been undertaken of the data that we have identified, our expectation is that existing observational data on the health effects of treatment cannot be used to quality adjust a general secondary care output index. In our memorandum to the Department (Appendix 1), we have recommended the systematic collection of health outcomes data from a sample of NHS patients. In implementing this recommendation, consideration needs to be given to the sample of NHS patients from which the information is to be derived, the choice of instrument, timing of measurement, the grouping of NHS activities, and the frequency of data collection. Our memorandum details these issues more comprehensively.

3.1.5 Expert Groups

A fourth source of health outcomes data would be through convening expert groups. Clinical expert could provide estimates of $\Delta v(t)$ without the need to deny cost-effective treatment to some patients for some treatments, as would be required if clinical trial data were used to update estimates of the effects of treatment. However, expert groups require information on which to make estimates of $\Delta v(t)$ and for many types of care where there have been no trials such data will be very sketchy. The snapshot before and after data which we have suggested should be gathered routinely would assist such expert groups so that in the future it should be possible to use them to estimate $\Delta v(t)$ for subsets of NHS activity.

The use of expert judgement is a well-recognised strategy for estimating health effects It can be used, for example, where the usual imperative for a randomised control study is resisted because of ethical considerations, or where the impact of health interventions can only be monitored over a lengthy time horizon, or where the natural incidence of a condition is low. In considering this approach and its potential for the present application, the following information is required :

- A description of the changes in health states for given conditions and for corresponding interventions;
- The valuation of those effects in terms of a standard metric (cardinal health status measure)
- The projection of those changes in health status over time (potentially over the entire life expectancy of patients with the particular condition), including any change in life expectancy.
- Estimation of the population prevalence of the condition / intervention.

Whilst a description of the natural history of a condition under different regimes should be within the competence of most clinicians, it is improbable that they will feel comfortable with the requirement that such description must be expressed in terms of a standard (generic) health status measure or index. This requirement is essential if health outcomes are to be aggregated across therapeutic areas and interventions.

The usefulness of different approaches to capturing expert judgement varies according to application. Such judgements may be elicited in a number of ways, from individuals acting in isolation or as a consequence of the deliberation of specially empanelled groups. Where the objective is to establish parameter estimates in a well-defined model then the input of a small number of experts acting in isolation can be productive, as exemplified by the derivation of profiles of patient health status constructed in the classic evaluation of CABG (Williams, 1985). This relied upon the capacity of experts (in this case cardiologists) to convert hypothesised patient scenarios into a standard generic classification of health status. Cardiologists were asked to indicate future health states that patients would be in given alternative forms of treatment for their heart condition.

One of the most ambitious projects in this field in recent times has been the Dutch attempt to establish disability weights for diseases as part of efforts to build a burden of disease model for the Netherlands (Stouthard *et al*., 1997). This enterprise was driven by the earlier example of the Global Burden of Disease project (World Bank, 1993; Murray and Lopez, 1996). A list of 52 diagnostic groups was selected on the basis of their importance to public health. It covered some 70% of all causes of death and some 65% of total health care costs. Descriptions of diagnostic conditions were made in terms of a standard set of dimensions based on EQ-5D as formulated by the study team. Person trade-off (PTO) and VAS rating methods were used to elicit weights for these conditions using 38 physicians divided into 3 expert panels. A fourth, lay panel was set up using well-educated individuals working outside the healthcare sector The weights on the dimensions were then applied to derive overall health status measures. Some important pointers can be taken from this Dutch study

- it is highly improbable that a single medical expert has real insights into the consequences of all diseases;
- even if such individuals exist, members of a panel will have complementary experience/expertise;
- diagnosis implies other information about prognosis, treatment, outcomes, potential to benefit;
- this implicit information is likely to be unevenly distributed across a panel;
- standardised descriptions of diagnostic conditions is helpful in reducing these problems.

The Dutch study indicates the significant resource commitment required to address just one part of the more general problem of valuing health benefits over time. There was no attempt to project the lifetime consequences of interventions, to contemplate the impact on mortality or to quantify gross changes to population health. It seems probable that efforts to synthesise domestic UK data in this field would require some time to develop, perhaps over a 2-3 year period. Comparable development initiatives in, for example, the derivation of a standard descriptive nomenclature for medical histories required both time and commitment from its expert panels. Success in such enterprises is rarely assured.

In addition to the resource implication, there are several key issues that highlight the limitations that could follow the use of assessments determined by expert groups. The first is the reliance on judgements arising from within the professional body most intimately concerned with the delivery of the care under review. Whilst it would be unfair to suggest that this proximity would necessarily compromise those judgements, it would be sometimes difficult to isolate the suggestion of vested interest. A self-policing approach is unlikely to sustain long-term public appreciation, especially if it is adopted as an alternative to the collection of more appropriate forms of data – for example, data that capture real as opposed to hypothetical outcomes data.

Second, where convergence amongst experts can be relatively easily achieved, the approach can inform models of health that, if not wholly accurate, are nonetheless reasonably robust. Where there are competing treatment modalities or where there is genuine uncertainty about future health trajectories the approach can be more problematic. For example, the development of consensus guidelines in respect of more diffuse health care practice can require significant investment in face:face meetings and is not always assured of a settled result.

Third, inevitably it would be costly to organise expert groups for all areas of NHS activity. A more limited coverage may render year on year comparisons problematic.

Despite such problems, there are advantages to the use of expert groups, the foremost of these being that they do not depend upon the collection of new data but rather are

based upon the collective experience of clinicians directly concerned with the delivery of care. who have an understanding of the condition, its treatment and management. Where relevant, such expert groups should be expected to take into account the views of patients and other non-professionals. Since the rate of technology change is uneven across the range of therapeutic conditions that face the NHS it would be possible to be selective in the commissioning of any expert-led assessments. The periodic reconvening of expert groups could be programmed to avoid unnecessary duplication of effort in those areas in which the pace of change seems likely to be slow. It may be best to utilise expert groups primarily for purposes of validation and triangulation of the other approaches to estimating the change in health outcomes secured by NHS interventions.

Given the resources available to the project team we will only consider the use of expert groups in connection with the two or three disease groups we will examine to triangulate other methods of estimating productivity growth.

3.2 Readmission rates

One measure of the quality of hospital care is the probability that a patient is readmitted for the same condition. Readmissions may reflect poor care, in particular premature discharge, possibly resulting from pressure to reduce length of stay (Hofer and Hayward, 1995; Ludke *et al.*, 1993; Thomas, 1996). If there were data on the time profiles of health status, data on readmissions would add little, if anything to the measurement of NHS outcomes and the quality of care.

Readmission rates at NHS Trust level have been published by the Department of Health for a number of years. For example a 30-day readmission rate has been published for 1995/96 to 1998/99. This was amended and calculated as the percentage of emergency readmissions within 28 days, all ages, age sex standardised and published from 1998/99 onwards. A few condition specific rates have been published since 1998/99: emergency readmissions following hip fracture within 28 days, and emergency readmissions within 28 days following stroke, both age and sex

standardised. More recently, since 2001/02 data have been published on the percentage of emergency readmissions following discharge within 28 days (and also 7 days), for children. These data have been used by the Department of Health, CHI and the Healthcare Commission, in the construction of the star ratings for NHS Trusts.

The published NHS Trust-level readmission rates (Department of Health, 2001; 2002a; Commission for Health Improvement, 2003; Healthcare Commission, 2004a) are only adjusted for differences in age structure and gender and do not take account of severity of illness, co-morbid conditions, complexity, length of stay, as well as variation in admission practice.

It is possible to use Hospital Episodes Statistics (HES) data to calculate readmission rates at NHS providers using the within HES patient identifier which uniquely identifies a patient across all data years after 1997/8. The HES identifier could be used to construct condition-specific readmission rates with some degree of risk-adjustment.

We will explore in Phase 3 the potential use of readmission rates as a means of 'triangulating' results on quality trends for a chosen number of conditions / HRGs.

3.3 Waiting times

The need to wait for diagnostic tests and treatment may affect individual utility in two ways. First, delay may be associated with deterioration in the patient's condition. The pain and distress while waiting for treatment results in a loss of quality adjusted life years for the patients affected. In extreme cases delay may result in premature death. The benefit of reducing waiting time could be measured as the change in expected QALYs associated with a given reduction in the time waiting for treatment. A recent survey of the literature on the cost of waiting (Hurst and Siciliani, 2003) found some evidence on deterioration and premature death associated with waiting for cardiology treatment but little for other procedures. Clinical reassessment of patients on a waiting list was thought to contribute to reduction in adverse outcomes of waiting but there are little data on the frequency or efficiency of re-classification of

patients on waiting lists. If the NHS begins the routine collection of data on health related quality of life, the QALY improvement due to reduced waiting time should be captured by trend changes in QALYs.

Second there is some evidence that people view waiting time as a disutility independent of the effect on health outcomes. In this case a direct valuation of reduced waits will be necessary. We will explore both of these approaches in Phase 3.

Ideally we want to measure changes in waiting time from GP referral to admission for treatment by procedure (or HRG). With present DH data this is not possible. We can measure waiting time from GP referral to first outpatient appointment (NHS Trust return QM08) at the specialty level but not by procedure or HRG. The DH has discontinued collection of data on waits for second or more outpatient appointments. Waiting time from the date a patient is added to the inpatient waiting list to the date of admission for treatment is available by procedure and HRG from Hospital Episode Statistics.

In the short term there is no choice but to use the restricted measure of waiting time generated from HES data: the period between admission to the inpatient waiting list and admission for treatment. Even with this limited measure there is the issue of whether changes in waiting time should be calculated with reference to the mean or median waiting time for patients in each HRG. Siciliani and Hurst (2004) show there is a very large difference between these two measures—the mean waiting time being roughly double the median waiting time for English acute Trusts. This reflects the influence of the relatively few patients with very long waiting times, some of whom may have died or decided to seek treatment in the private sector. The median is likely to be the more appropriate measure. In Phase three we will examine the robustness of productivity growth measures to choice between use of median and mean values.

3.4 Choice and certainty of date of treatment

Over the next few years it is expected that an important change in the quality of NHS services will be an increase in booked admissions and a reduction in cancelled

operations as elective and emergency care is separated. Patient surveys and evidence from purchase of private health care suggest that patients value the ability to choose the date and time of treatment. Hospital Episode Statistics contain a field for method of admission, elective-booked. This information has been collected since 1989 and it will be possible to monitor over time changes in the proportion of patients treated in NHS Trusts that were able to choose and book admission date.

Increased use of Independent Treatment Centres (ITCs) may reduce the value of this measure of quality change. It is not clear what data ITCs will be required to submit to the DH. It is reasonable to assume a very high proportion (all?) ITC activity will be booked but we also need to know their activity levels by procedure and HRG if it is to be aggregated with that of NHS Trusts. The growth of NHS Treatment Centres does not create the same problem. Activity in NHS TCs is routinely reported in the returns of the Trust to which the NHS TC is affiliated.

With existing data it will not be possible to monitor changes in cancelled operations. We have figures for cancellations *before* the patient arrives at the hospital but not for cancellations *after* the patient has been admitted. Unless the DH or one of the regulators decides the issue is of sufficient importance to require regular returns, this aspect of performance cannot be included in our quality measure.

3.5 Environment

One characteristic of hospital output that is gaining increasing importance is hospital cleanliness, and the control of infections. A number of different data sources exist measuring these attributes. Cleanliness is given a great deal of attention in performance assessment of Trusts and is one of the key targets in the star ratings of NHS Trusts. The implicit assumption is that better cleanliness will lead to a reduction in hospital acquired and other infections, though infections are also now specifically being measured by various surveillance systems.

3.5.1 Cleanliness

Patient Environment Action Teams (PEAT) inspect hospitals and report on the patient

environment (cleanliness and tidiness) and food services. A weighting process is applied which emphasises the areas that patients deem most important. This includes:

- Cleanliness in wards
- Cleanliness in emergency departments
- Cleanliness in other departments and waiting areas
- Cleanliness of toilets in wards
- Cleanliness of toilets in emergency departments
- Cleanliness of toilets in other departments, and waiting areas
- Cleanliness of bathrooms in wards
- Environment in toilets in wards
- Environment in toilets in emergency departments
- Environment in toilets in other departments and waiting areas
- Environment in bathrooms in wards

3.5.2 Infection control

A number of different surveillance systems have been in place (both voluntary and mandatory) to monitor infections in English hospitals. These fall into three categories: wound infection, overall measures of infection control, and MRSA (hospital acquired infection).

In 1996, the Department of Health together with the Public Health Laboratory Service (PHLS), now incorporated in the Health Protection Agency, established and launched the Nosocomial Infection National Surveillance Scheme (NINSS). The NINSS scheme ran until November 2002, on a non compulsory basis. Further, it was a selective scheme in that each participating hospital could select one or more categories for surgical procedures to monitor among the list of 12 designated procedures. It was an intermittent scheme, in that hospitals were required to collect data for a minimum of three consecutive months each year. Some hospitals decided to collect the data continuously. A summary of the data collected is reported in 'Surveillance of Surgical Site Infections in English Hospitals 1997 – 2002' (Health Protection Agency, 2003).

In October 2000 the Department of Health proposed to introduce compulsory monitoring of hospital acquired infections (referred to as healthcare associated infections) in all NHS Trusts in England. This monitoring would extend to certain blood stream infections (including MRSA), as well as infections of wounds following orthopaedic surgery. The orthopaedic SSI surveillance is compulsory as from 1st April 2004, while the MRSA surveillance scheme has been in place since 2001. (Department of Health, 2003).

3.5.2.1 Wound infections

Wound infections can occur as a consequence of many surgical procedures. It appears that 50 to 70 per cent of all wound infections occur after patients are discharged from hospital (Department of Health, 2002b). These types of infections are of concern to secondary, primary and community care providers within the NHS.

A new surveillance scheme for surgical site infection (SSI) has been introduced in April 2004. It is again a voluntary scheme, although the data collection on wound infections occurring after orthopaedic procedures is mandatory for all NHS Trusts in England. (Health Protection Agency, 2004).

3.5.2.2 MRSA

Methicillin resistant Staphylococcus aureus, better known as MRSA, is a blood stream infection, which has been increasingly affecting NHS patients in England (Crowcroft and Catchpole, 2002; Health Protection Agency and Office for National Statistics, 2004). A mandatory Department of Health Bacteraemia Surveillance Scheme for MRSA has been in place since 2001. Data are available by NHS Trust for the first three years of its existence (Department of Health, 2004b). These data include the number of MRSA bacteraemia reports by Trust and the MRSA rate per 1000 bed-days by Trust from 2001/02 to 2003/04. Data are disaggregated by Trust type: General Acute, Single Specialty and Specialist Trusts.

At present, data on wound infection and infection control are not comprehensive and

and cover a relatively short period of time. Over the longer term this may prove a useful quality indicator but for our work on quality change over the next nine months it is unlikely we can do much with existing data.

3.6 Patient satisfaction: Inpatient Survey 2004

It is possible to measure the non-health related characteristics of the health sector using information from patient surveys. The Healthcare Commission carried out five national surveys asking patients across England about their experiences of different NHS services[4]. A total of 568 NHS organisations and 312,348 patients took part in the surveys. Each trust that took part identified 850 eligible people. Patients were sent questionnaires, and up to two reminders. Response rates to the surveys varied from 63% for the adult inpatient survey to 42% for the mental health survey.

Over 88,000 patients were involved in the *Inpatient Patients Survey 2004*. The survey asked patients from 169 acute and specialist NHS trusts across England about their recent experience of inpatient care. Patients were eligible to take part if they had had at least one overnight stay, were over 18 years old and were not maternity or psychiatry patients. Completed questionnaires were returned by similar proportions of men and women. Response rates were highest for 51-81 year olds, and lowest for 18-35 year olds.

Such surveys contain a great deal of information on many different aspects of the patients' experience. The primary goal, therefore, is one of data-reduction whilst minimising loss of information. In the case of a dataset containing continuous variables, one could use factor analysis (e.g. Harman, 1976). However, the information contained in the inpatient survey is almost exclusively categorical in nature. Patients are asked to rank their responses. For example, in answer to the question 'How would you rate the hospital food?' they are given four ranked choices from 'poor' to 'very good.' In such cases one cannot use factor analysis of raw data;

[4] Two of the surveys, the adult inpatient and primary care services surveys, have been carried out before. The first inpatient survey was conducted in 2002.

one must first estimate polychoric correlation in order to extract commonalities. There is, however, a further complication since we do not have access to the primal information (i.e. the information at the level of individual patient responses). Instead, we have summary statistics of the proportion of patients responding to each category of these questions in each hospital (e.g. 10% answering 'poor', 20% answering 'very good' etc.). It is possible to use a method similar to Carlson and Parkin (1975), using the proportions to identify the thresholds in the underlying variable. This can be extended to account for the fact that we have latent, rather than observed variables. This is possible if we assume that the indicated thresholds are as close as possible to the population average. It is also possible to relax the assumption of normality of the underlying latent variable (Mitchell, 2002). The means of the latent variable for each hospital would then be used in a standard factor analysis to obtain summary measures of patient satisfaction.

3.7 Weights for activities and characteristics

A central problem in creating a single index of output is identification of weights that can be used to sum diverse outputs and their characteristics. In theory the weights should reflect the marginal social value of the outputs. The problem is identifying data relevant to estimate or approximate social value in the absence of market prices.

3.7.1 Expenditure weights

3.7.1.1 Introduction

As we argued in our First Interim Report (section 2.10.3) under certain assumptions, marginal costs reflect marginal valuations. In such situations, a case can be made for using marginal costs as a basis for determining the relative values of different activities. Current NHS practice is to use unit (average) costs derived from the Reference Costs (Department of Health, 2004c). In an attempt to test whether the use of marginal costs would make a difference to calculated productivity growth we will explore whether it is possible to derive estimates of marginal costs and compare them with reference costs. We have made some highly preliminary estimates of marginal

costs and will continue the analysis in stage 3 of the project.

3.7.1.2 Methods

We will compare average cost estimates derived from accounting data and marginal cost estimates derived from a regression model.

Accounting Estimates. Collected annually from all NHS Trusts, the Reference Costs comprise average costs for different types of activity, defined predominantly by Healthcare Resource Group. These average costs are accounting costs arrived at by applying national guidance about how to apportion shared input costs to each activity type.

In our preliminary analysis we calculated a national Reference Cost - $\overline{C}_{jt}^{RC}$ - for each elective HRG j at time t as a weighted average of the Reference Costs reported for elective activity conducted in inpatient and daycase settings. The weights represent the share of activity (FCEs) in the two settings for each HRG. Hence:

$$(1) \qquad \overline{C}_{jt}^{RC} = p_{jt}^{E}(\overline{R}_{jt}^{E}) + (1 - p_{jt}^{E})(\overline{R}_{jt}^{D})$$

where $\overline{R}_{jt}^{E}$ and $\overline{R}_{jt}^{D}$ are respectively the mean elective and daycase Reference Costs reported by Trusts in HRG j and $p_{jt}^{E} = FCE_{jt}^{E} \Big/ \left[FCE_{jt}^{E} + FCE_{jt}^{D}\right]$ where FCE_{jt}^{E} and FCE_{jt}^{D} are respectively the number of elective and daycase FCEs nationally in HRG j. This is consistent with the methodologically used to calculate national tariffs for the Payment by Results reform.

The annual figures were summarised into a single estimate covering the full period for which data were available, with the annual activity in each HRG used to weight the annual mean costs.

Regression Estimates. The alternative approach to calculation of marginal costs is to

derive them from a regression in which total costs are regressed on the amount of activity undertaken in each output activity. The parameters from this regression can be interpreted as estimates of marginal cost. Hence, we estimated a regression model of the form:

$$TC_{it} = \alpha + \beta_{1t} x_{1it} + \beta_{2t} x_{2it} + ... + \beta_{Jt} x_{Jit} + e_{it} \quad (2)$$

where TC_{it} is the total cost of all types of activity in Trust *i* at time *t*. This is calculated by multiplying the number of FCEs by the reference cost for those FCEs in each HRG and summing across all HRGs. β_{1t} can be interpreted as the marginal cost of treating an additional patient in HRG 1 at time *t*. We ran regressions for each year separately, but also created a panel dataset combining all six years' worth of data.

We tested functional form by running linear models and log-log models. We tested the inclusion of a size variable namely, the average number of beds in the Trust and various non-linear combinations of this variable. We also tested the inclusion of time dummy variables. We ran a number of different estimation procedures including pooled OLS, random and fixed effects models.

The objective is to compare $\overline{C}_{jt}^{RC}$ with estimates of β_{jt} from the various specifications of the regression model.

3.7.1.3 Data

We used six available years of Reference Cost data from 1998/99 to 2003/04. These give, by NHS Trust, the number of FCEs, and the average cost, in each HRG for elective, daycase and non-elective (emergency) activity respectively. We generated a single dataset that combined elective and daycase activity and costs. We excluded PCTs (some of which began to provide acute care from 2001/02) since their production and cost structure may differ from those of NHS hospital Trusts.

In the regression model, we would wish to use the number of FCEs in each HRG as

an explanatory variable. This however gives rise to serious degrees of freedom problems with over 500 HRGs and thus RHS regressors. To address this limitation, we selected only large volume HRGs for inclusion in the model, ensuring that at least one HRG from each HRG chapter is represented. We then created composite HRG categories for the smaller volume HRGs within each HRG chapter. All 98 codes (chemotherapy) and 99 codes (complex elderly) across HRG chapters were added together into 2 composites. Volumes were determined by running frequencies for all HRGs in 2002/03. This exercise resulted in a total of 35 individual HRGs being included in the model, together with eighteen composite HRG chapter variables.

3.7.1.4 Discussion

This exercise produced some significant and plausible coefficient estimates, but also a few coefficients with negative signs suggesting a negative marginal cost. Possible reasons for this may be:

1. issues of scale economies
2. misspecification in the model
3. misallocation of costs between elective and daycases and non-electives

We intend to carry out further analysis in Stage 3 of the project to arrive at a view as to whether marginal costs differ markedly from average costs and hence whether it is worthwhile for the DH to investigate whether marginal cost estimates for HRGs could be derived from the raw accounting data used to estimate reference costs.

3.7.2 International prices

3.7.2.1 Introduction

There is a precedent in cost benefit analysis for using world prices to value domestic output when domestic prices are absent or distorted. The rationale is that because trade could take place at world prices, they are legitimate measures of opportunity cost to the domestic economy.

This rationale has limited force in the context of the NHS because there is not a significant world market in health care. Moreover a free market in health care exists in

no country. Any domestic price or cost information that is available will not be a reflection of competitive conditions but tend to be administered prices subject to stringent domestic regulation or negotiation.

These caveats notwithstanding, we have explored whether the valuations of activity would be sensitive to the use of price information from other countries.

3.7.2.2 Data

While many countries may well collate price or cost information about health care interventions, this information is rarely placed in the public domain. Data similar to the English reference costs are posted on the internet in Australia, New Zealand and Italy.

We compared the data for Australia and Italy to the reference cost data for England. Data for New Zealand were not considered further because, except for minor differences, these are identical to the data for Australia.

3.7.2.3 Derivation and mapping of classification systems

An obvious problem in using price information from other countries is that 'product' definitions may differ. The three countries considered here have developed their own systems to classify health care activity. There are a number of reasons for the emergence of different classification systems, among them the following:

- The use of different versions of the International Statistical Classification of Diseases and Related Health Problems (ICD) coding system. Version 10 is used in the UK and Australia; version 9-CM in Italy.
- The use of different systems to code procedures;
- Differences in the extent to which the medical profession is involved in devising the classification system;
- Differences in the methods used to develop the system, including the form of statistical model, variables used, and the quality of the underlying data;
- Differences in medical practice.

Thus in comparing price information the first challenge is to ensure that prices apply to the same activities.

Even for acute sector activity, the product list in each country is extensive (565 version 3.1 HRGs in England; 588 DRGs in Italy; 661 AR-DRGs in Australia) so a complete mapping was impractical for present purposes. Instead we limit our comparison to a handful of HRGs, these being:

- Cardiac valve procedures;
- Coronary bypass;
- PTCA;
- Arthroscopy;
- Bilateral primary and primary hip replacement;
- Bilateral primary and primary knee replacement.

Table 3.8 details the mapping of these HRGs to counterpart groupings in the Australian and Italian classification systems. In general for these procedures the Australian and Italian systems are more delineated than that of the UK. For example, cardiac valve procedures form a single HRG category in the UK (E03). This HRG comprises 48 separate OPCS procedure codes so may well be a fairly heterogeneous set of activities. In contrast to the UK's single category, in Australia cardiac valve procedures are amalgamated into three categories, with patients distinguished according to whether or not a pump was used and whether or not patients suffered complications and comorbidities. In Italy a distinction is made according to whether or not the patient was catheterized.

In each country, more than one category is available to describe hip replacements. However, the basis of sorting patients into one or other category differs. In the UK, those having bilateral hip replacements are distinguished from others. In both Australia and Italy, patients are distinguished according to whether or not they suffered complications and comorbidities, with age also being used to categorise patients in Italy. To achieve some degree of comparability, therefore, the sub-

categories in each country have been amalgamated into a global category including all hip replacements. Similar arguments apply to knee replacements.

Table 3.8

England HRG Description	Code	Australia DRG Description	Code	Italy DRG Description	Code
Cardiac valve procedures	E03	Cardiac Valve Proc w pump invasive cardiac Inves procedure	F03Z	Interventi sulle valvole cardiache con cateterismo cardiaco	104
		Cardiac Valve Proc w/o pump invasive cardiac Inves procedure w cat or sev cc	F04A	Interventi sulle valvole cardiache senza cateterismo cardiaco	105
		Cardiac Valve Proc w/o pump invasive cardiac Inves procedure w/o cat or sev cc	F04B		
Coronary bypass	E04	Corony Bypass W Invasive Cardiac Inves Procedure W Catastrophic CC	F05A	Bypass coronarico con cateterismo cardiaco	106
		Corony Bypass W Invasive Cardiac Inves Procedure W/O Catastrophic CC	F05B	Bypass coronarico senza cateterismo cardiaco	107
		Corony Bypass W/O Invasive Cardiac Inves Procedure W Catastrophic CC	F06A		
		Corony Bypass W/O Invasive Cardiac Inves Procedure W/O Catastrophic CC	F06B		
Percutaneous transluminal coronary angioplasty	E15	Percutaneous Coronary Angioplasty W AMI	F10Z	Interventi sul sistema cardiovascolare per via percutanea	112
		Percutaneous Coronary Angioplasty W/O AMI W Stent Implementation	F15Z		
		Percutaneous Coronary Angioplasty W/O AMI W/O Stent Implementation	F16Z		
Bilateral primary hip replacement	H01	Hip Replacement W Cat or Severe CC or Hip Revision W/O Cat or Severe CC	I03B	Interventi su anca e femore, eccetto articolazioni maggiori, eta >17 con CC	210
Primary hip replacement	H02	Hip Replacement W/O Cat or Severe CC	I03C	Interventi su anca e femore, eccetto articolazioni maggiori, eta >17 senza CC	211
				Interventi su anca e femore eccetto articolazioni maggiori, eta <18	212
Bilateral primary knee replacement	H03	Knee Replacement and Reattachment W catastrophic CC	I0A4	Interventi su ginoccio con CC	221
Primary knee replacement	H04	Knee Replacement and Reattachment W/O catastrophic CC	I04B	Interventi su ginoccio senza CC	222
Arthroscopies	H10	Arthroscopy	I24Z	Artroscopia	232

3.7.2.4 Price information

Price information for England is taken from the reference cost database for 2003/04, compiled from the mandatory costing returns made by all English Trusts. For the comparative purposes of the present exercise the HRG price is calculated as the mean weighted average of elective inpatient and daycase activity, this being consistent with the method used to calculate the national tariff introduced under the Payment by Results regime.

The Australia 'price' data are derived from the National Hospital Cost Data Collection (http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/health-casemix-costing-costmain1.htm). The NHCDC is an annual, voluntary collection of hospital's cost and activity data. It represents 75% of activity from public hospitals.

We analysed data for 2001/02.

The Italian Ministero della Salute (former Ministero della Sanità), which is the equivalent of the English Department of Health, is responsible for setting the criteria to determine the tariffs for the reimbursement of a range of healthcare services, including hospital services. The tariffs represent the maximum amount that can be reimbursed to healthcare providers, both in the public and in the private sector. The latest criteria were set out in the Decreto Ministeriale 15/04/1994. (D.M) (Departmental/Ministerial decree).

The 19 Italian regions and 2 autonomous provinces are responsible for calculating the tariffs to be applied within their territorial area, usually basing calculations on costs in a sample of their local public and private health care providers. Consequently, there is price variation across regions, with the correlation among regional tariffs ranging from r=0.80 to r=1.00. The Ministero della Salute uses this information to set national tariffs, which apply only if regions fail to undertake their own calculations. The tariffs established in 1997 still appear to apply and the national tariffs were used in this exercise. Tariffs are available from the following website: http://www.ministerosalute.it/programmazione/resources/documenti/all97euro.XLS

Where categories are an amalgam of Australian or Italian DRGs, an overall 'price' is calculated as the arithmetic mean of the separately reported DRG prices. No further data (on the volume of activity, say) were available to enable alternative calculations.

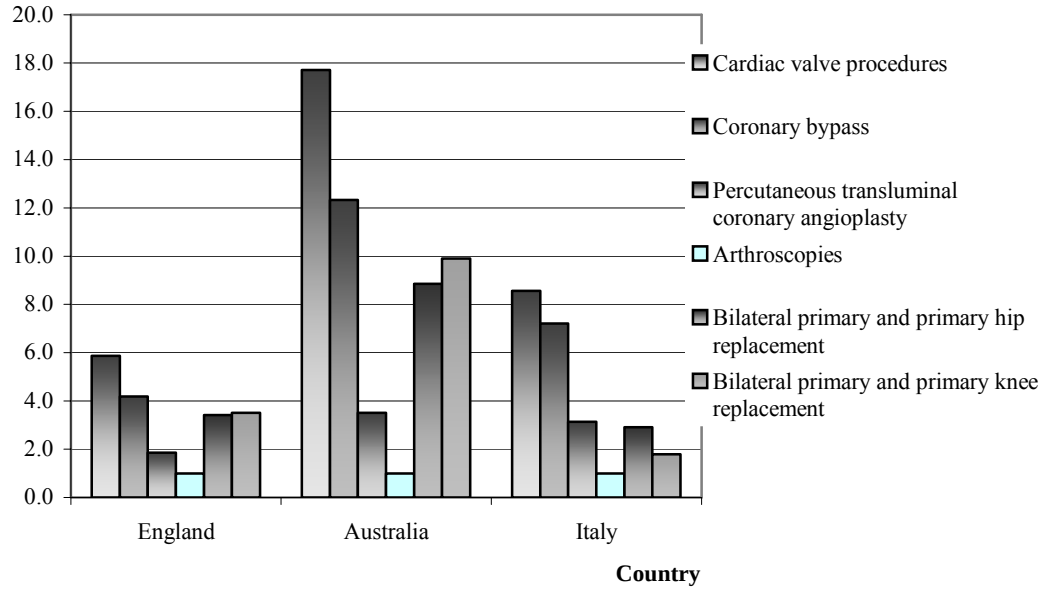
3.7.2.5 Results

Table 3.9 shows the tariffs/costs converted into sterling at current exchange rates for the six HRGs, together with their weight relative to the price of an arthroscopy in each country. The relative values are plotted in Figure 3.7.

Table 3.9

Intervention	England		Australia		Italy	
	£	weight	£	weight	£	weight
Cardiac valve procedures	13836	5.9	12247	17.7	11781	8.6
Coronary bypass	9863	4.2	8530	12.3	9932	7.2
Percutaneous transluminal coronary angioplasty	4357	1.8	2433	3.5	4318	3.1
Arthroscopies	2356	1.0	692	1.0	1377	1.0
Bilateral primary and primary hip replacement	8045.5	3.4	6128	8.9	4023	2.9
Bilateral primary and primary knee replacement	8264	3.5	6853	9.9	2472	1.8

Figure 3.7 - International prices for 6 HRGs: comparisons by country, rebased around arthroscopy



The English and Australian rankings of these activities are identical, with the difference in relative values driven only by the low cost of arthroscopies in Australia. There is a lower correlation (r=0.84) between the English and Italian relative values, this being a reflection of the lower costs of the three orthopaedic interventions. This implies that the valuation of these activities would be more sensitive to the use of Italian than Australian data. However, it would be unwise to extrapolate from this handful of activities.

3.7.2.6 Conclusions

Use of price or cost information collated by other countries to value NHS activities or outputs is inadvisable for three main reasons:

1. There is no world market for health care and all domestic prices are distorted in some way.
2. There is no common terminology to describe health care activities.
3. The valuation basis appears sensitive to which country the data come from.

3.7.3 Valuing other characteristics

In addition to health outcomes, we have identified a number of other valued characteristics of health care. These include waiting times, patient satisfaction, choice of treatment date, and the environment in which care is delivered.

In our First Interim Report we examined available evidence on valuation of changes in waiting times derived from willingness to pay and discrete choice experiments. No additional relevant experimental data have been identified. Valuations derived from discrete choice experiments are infrequent and rarely seek to elicit population values as opposed to valuations of patients with particular conditions (Bryan and Dolan, 2004). Given the paucity of evidence from these studies, valuations from discrete choice experiments are unlikely to be suitable for the generation of weights to be used in an index of output intended to routinely monitor, on at least an annual basis, changes in NHS output. At best the data can be used in a sensitivity analysis of changes in the mix of non-priced outputs.

It may be possible to apply QALY valuations to reductions in waiting times. This is justifiable if one can consider a reduction in waiting time as translating into earlier receipt of health improvements, leading to a change in total QALYs obtained. The marginal contribution to the total QALY amount brought about by reduced waiting time can be valued by applying a monetary value for a QALY.

No known data have been identified that might be used to place values on the other characteristics of health care outputs. Instead we will explore the use of incorporating information about these characteristics as output scalars. We shall explore the sensitivity of the productivity measure to alternative assumptions about the weights attached to these scalars.

4 Measurement of inputs

This section considers the measurement of inputs. As in previous reports inputs are divided into three broad groups, labour, intermediate and capital. In addition the note also considers inputs whose use contributes to producing future rather than current output, namely R&D expenditure and training costs. Growth in the three inputs will be combined into an aggregate index, using expenditure shares as weights, and then deducted from output growth to yield estimates of productivity growth. However it is useful to track changes in the three inputs separately, and in their components, to see to what extent changes in output reflect differing patterns of input use and substitution between inputs. Therefore the data required are discussed with this general model in mind.

For each of these four categories we first set out the methodological framework so that this can be readily linked to data requirements. We then discuss data identified and those that can be used in the next nine months. We then discuss what we plan to do with the data that differs from existing DH/ONS approaches and finally list data DH should start to collect.

4.1 Labour

4.1.1 Measurement and data requirements

The standard approach to measuring labour input is to use number of hours worked, differentiated by type of worker, e.g. as recommended in the OECD's manual on *Measuring productivity*, OECD (2001). The simplest measure is a headcount of

persons engaged, including self-employed. However this will hide changes in actual hours worked due to changes in part-time working, changes in holiday entitlements or other absence from work and shifts in normal working hours that can occur due to legislation, e.g. the junior doctors working time directive. Thus ideally data on actual hours worked are required. A half-way measure between a simple headcount and actual hours is full-time equivalent persons engaged which attempts to correct for the extent of part-time working. These measures however frequently employ a crude adjustment for part-time working, such as counting a part-timer as equivalent to one half of a full-time person. Thus changes such as reduced hours by junior doctors will not be captured. It is also useful to draw a distinction between paid and actual hours worked. The former are generally based on contractual normal hours whereas the latter includes hours worked but not paid (unpaid overtime) and excludes time paid but not worked due to holidays, sick leave etc. Ideally we require information on actual hours worked but this is the most difficult concept to measure. Finally in measuring labour input for a specific sector, such as the NHS, it is important to include all hours spent working in that sector even if the individual is technically employed by another sector. Thus in the NHS context, it is important that agency staff are correctly allocated.

Differentiation of labour by type of worker is required to take account of the fact that effort and skills of persons vary by type. As the OECD manual points out, *"Because a worker's contribution to the production process consists of his/her "raw" labour (or physical presence) and services from his/her human capital, one hour worked by one person does not constitute the same amount of labour input as one hour worked by another person*", OECD (2001, p. 41). Volume measures such as headcounts or total hours treat all workers equally but the output produced by a senior consultant is very different from a junior doctor or nurse. Thus a quality adjusted labour input measure is required that takes account of the productivity impacts of different types of workers.

A standard approach to measuring quality adjusted labour input is to divide the labour force by type of worker and then employ the shares of each type in total labour compensation as quality adjustment factors. Growth in the volume of labour input is the growth in each type weighted by their shares in total (across all types) hours

worked. Quality adjusted labour input growth instead weights growth in hours for each type by their shares of total labour compensation, thus giving a higher share to more highly skilled staff. Implicit in this calculation is the assumption that unit labour compensation reflects the marginal productivity of each type of worker, an assumption that may not be correct in practice if input markets are not perfectly competitive.

Unit labour compensation is the price of labour input. Data on labour compensation are required for two reasons. The first, outlined in the previous paragraph is as a measure of the output contributions of workers of different types. The second use however is the need to weight labour input with other inputs such as capital and intermediate inputs to derive an aggregate input measure. Total expenditures on labour can be used to derive labour input's weight but it is important to distinguish payments for labour services from other returns to the individual such as a return to capital for the self-employed.

4.1.2 Data identified

Headcounts by type of worker and whole-time equivalents are available from the *NHS Employment Census*. This gives detail by types of doctors, nurses, allied health professionals, other scientific and technical staff, health care assistants, managers and administrators, ambulance staff and other staff. Data are available annually from 1995 and contain considerable detail by type of staff for NHS and primary care Trusts. Numbers of GPs and staff employed in GP surgeries are also available from this source.

In addition the annual *labour force survey* (LFS) contains data on persons employed and actual hours worked by industry SIC code, and distinguishes those employed by the private and public sectors. Additional data on hours worked can be taken from the Annual Survey of Hours and Earnings (ASHE) (previously the New Earnings Survey). The LFS contains some information on qualifications of workers (e.g. university degrees, intermediate or low qualifications) although the sample size is likely to be small.

Expenditures on labour input are available from Trust financial returns although the level of detail by type of person employed is not as great as in the NHS employment census. Details on average pay rates by type of worker are available from the NHS annual earnings survey, for 1998/99, 2000 and 2002. These data are not as detailed as the employment data but we have requested DH to attempt some matching from unpublished data. Information on hourly pay rates are also available from the LFS and ASHE.

Estimating the remuneration of GPs will pose some difficulties as it is important to exclude earnings that represent a return on capital. Data from Inland Revenue Inquiries are likely to be useful in this respect. Also we could look at data from the doctor's pay review bodies as used by ONS (see below).

Using a combination of the above sources should yield a reliable quality adjusted measure of real labour input. Nevertheless, some complications are likely to arise that will not be adequately covered by existing data sources, for example capturing the impact of the junior doctors hours directive or the allocation of agency staff. In such cases, and given the time frame of the project, it may be necessary to resort to some sensitivity analyses by bounding the likely impact of these changes.

4.2 Intermediate inputs

4.2.1 Measurement and data requirements

Intermediate inputs are those purchased goods and services that are used up within a year. They include drugs, other clinical supplies, energy and fuel, other purchases from manufacturing and bought in services. Volume measures of intermediate inputs are typically measured by expenditures on these items deflated by appropriate price indexes.

It is useful to divide intermediate inputs into as many categories as feasible so as to

examine substitution within this group of inputs and with other inputs such as labour and capital. For example the trend towards outsourcing of non-health tasks, such as cleaning, by the NHS involves the substitution of purchased services for labour and purchased goods (supplies). Of more consequence is the issue of industry attribution of productivity change, in particular related to the use of pharmaceuticals. Expenditure deflated by a price index based on the cost of items matched through time will yield an adequate volume index for drugs. However, if new drugs appear on the market that lead to more favourable health outcomes then a correct price index should allow for these changes. Thus in the case where a new, improved but more expensive drug is used by the NHS to treat the same condition, a quality adjusted price index will show less price growth than implied by comparing the unit costs of old and new drugs. This will raise the (quality adjusted) volume of intermediate inputs and hence lower measured productivity. Without such an adjustment, part of the productivity change will be attributed to the health sector when it should instead be attributed to the pharmaceutical sector. The quality of other intermediate inputs such as clinical supplies may also change and so may require quality adjustments to their deflators.

A number of methods exist to quality adjust inputs, including hedonic methods; see Triplett (2004) for an extensive discussion. The hedonic method involves regressing price change on a number of characteristics of the product; in the case of drugs these would be primarily health outcomes. However the hedonic method is based on the assumption of competitive markets and is less useful if this assumption does not hold. In the specific case of the NHS, however, hospital drug prices are negotiated and new drugs are often initially sold at a discount in order to induce their use by health professionals. Alternatives would involve incorporating some assumption on the effectiveness of new drugs relative to either the drugs they are replacing or the average price of drugs currently employed to treat the same condition. Clinical expertise would be useful in this regard but obtaining such information is expensive and time consuming. Thus it will prove very difficult to incorporate a quality adjustment for all intermediate inputs, but case studies for one or two cases may be attempted. This should give some idea of the importance of quality adjustments.

4.2.2 Data identified

Expenditures on intermediate inputs by category for hospital and community health services are available in the Trust financial returns. These can be deflated by the Health Services Cost Index (HSCI) calculated by the Department of Health. In this project we will consider each item of intermediate expenditure and its deflator rather than take single values from DH. This way we can experiment with the use of alternative deflators and hopefully get a handle on the quality change issue. For example we could consider list prices in the IMS database for drugs, subject to DH delivering these prices to us, to see if we can match old and new drugs for particular conditions. We could then see the difference between assuming all the price change is quality improvement versus the one where the entire difference is assumed to be a price change.

Intermediate expenditures by GPs can be taken from Inland Revenue Enquiries, with Family Health services drugs from the Prescription Analysis and Cost (PACT) database.

4.3 Capital

4.3.1 Measurement and data requirements

Capital investments need to be treated differently from other inputs since assets purchased at any one time are used to produce output over periods greater than one year. If assets were rented by firms then the rental value equals the flow of productive services. Problems arise when assets are owned by the user so that market rental rates are not observed. In that case the quantity of capital services and their prices – user costs or rental prices – have to be imputed. The growth in the volume of stocks of a single type of asset is measured by means of the perpetual inventory method which cumulates investment and deducts retirements and allowance for decay or efficiency decline as the asset ages, i.e. any loss of productive services from the stock through time. Capital services are defined as the flow of productive services from the cumulative stock of past investments. The flow of services from any asset in any one

time period is generally assumed to be proportional to the stock. In the past these measures were aggregated by simply summing across asset types, equivalent to weighting growth rates using asset acquisition or market prices. These measures have been superseded in recent years by the calculation of volume indexes of capital services, which are (slowly) being adopted in national accounts. The new measures aggregate by using user costs or rental prices rather than market prices. Under competitive conditions user costs reflect the marginal productivity of the various assets. Employing user costs as aggregation weights is a way of incorporating the productive contribution of heterogeneous assets. User costs represent the amount of rent that would have been charged for one unit of an asset and consists of three terms. These are the cost of financing the asset (measured by the rate of return), the value of depreciation (including both physical decay and the fact that the asset's service life has declined by one year), and capital gains or losses.

Measuring capital services therefore requires data on investment in current prices, asset price deflators which adjust for quality change, assumptions on the age efficiency profiles of the assets and calculations of user costs or rental prices. Assets that depreciate rapidly have higher user costs than longer lived assets. In recent years whole economy investment had proceeded much more rapidly in short lived assets such as computing equipment so that the growth in the volume index of capital services has been higher than that implied by a simple sum of capital stocks. This finding is also likely to hold for the NHS, and is likely to be reinforced by the increased use of 'high tech' medical equipment.

4.3.2 Data identified

For capital input we need investment data by asset type in current and constant prices and/or estimates of the values of capital stocks for a benchmark year. For HCHS current price investment, the usual division into structures, machinery, transport equipment and information technology is available from NHS estates and from Trust financial returns. There are questions relating to the time period for which information is available. Long time series are required to estimate capital stocks, in particular buildings stocks. NHS estates also have information on the age of buildings, which can be used to as a check on the time series information and as an indicator of asset

lives/depreciation rates and may be useful for constructing initial stocks of buildings. The TFR also have some information on purchases of medical equipment, X-ray equipment and laboratory equipment as well as data on other assets.

Some information is available on investment through PFI, the date the tender started and time to completion. While NHS estates suggest that most PFI under short contracts is construction, they acknowledge that it is not possible to tell how much is shorter lived assets such as medical equipment without examining each contract. ONS suggest some PFI contracts cover much longer periods, up to twenty years and involve intermediate services as well as investment. It may be necessary to use ONS data to backdate NHS estates data. ONS calculate sector specific asset deflators which will also be used in this project.

It is unclear to what extent the data to be supplied by NHS estates include software. The most recent revision to the system of national accounts (SNA93) recommended that software be included in capital rather than intermediate purchases. Software can be estimated using the ONS Supply Use Tables. It might be useful to see if we can get information on software purchases directly from DH. Deflating investment in medical equipment by ONS deflators may not account for all quality improvements from high technology equipment, although ONS do appear to incorporate some quality adjustments in their producer price indexes. For this we would need more detailed information on types of equipment used and quality adjusted deflators. The latter are unlikely to be available but we could attempt sensitivity analyses using international deflators if available. Such an approach was used for computing equipment for EU countries in O'Mahony and Van Ark (2003). Alternatively we could request asset valuations for medical equipment and age profiles, similar to that available for buildings from NHS estates.

4.4 ONS/DH estimates

In their recent measures of productivity in the NHS, ONS measured labour input by expenditures on labour deflated by an index of wage rates. Their reason for using this somewhat unusual approach was concern over the quality of the headcount measures,

in particular the allocation of agency staff, and difficulties in measuring hours worked. For hospital and community health services, the deflator was the pay cost index (PCI) provided by the Department of Health. The deflator for general medical services and personal medical services was the net remuneration index from the Review body on Doctors' and Dentists' remuneration.

For the hospital and community service sector ONS used total intermediate expenditure deflated by the HSCI. Deflators for family health services were also derived from DH. Two variants were tried for family health services drugs (net of receipts from prescription charges) based on a Paasche price index for existing items or the average unit cost of all items. These two variants lead to quite big differences, amounting to about half a percentage point per annum from 1995 to 2003 in real input growth. ONS rely on the Prescription Pricing Authority (PPA) to carry out this adjustment rather than a detailed examination of the items included in PACT. ONS plan to consider the division by item in the HSCI in more detail and acknowledge that drug prices are a particular problem. In terms of intermediate input in general, and drugs in particular, there will be some overlap between this project and research carried out at ONS.

In their recent measures of NHS productivity, ONS use deflated capital consumption since this was the measure most readily available. However ONS acknowledge that this concept is not the theoretically correct measure since capital consumption is a concept related to wealth stocks rather than the productivity capacity of capital. Given this problem ONS also include a variant of capital input based on national accounts estimates of capital services for a broader sector than the NHS, the health and social services sector. ONS intend to pursue capital services measurement further.

There will be close collaboration with ONS on measuring inputs throughout the project. ONS will continue to refine their expenditure deflated measure of labour input, and is interested in the development of direct volume measures for the NHS. Meetings will be arranged to compare results from different approaches with the suggestion that ONS might move to an hours based measure if they are convinced on its reliability. On intermediate input, it was agreed that simultaneous research by both the NIESR/CHE and ONS teams was potentially useful and that the teams should

meet regularly to compare progress. It was also agreed that the measurement of capital services should be a joint collaborative effort of ONS and NIESR, especially since NIESR are heavily involved in measuring capital input for private market sectors of the economy. It was agreed that we would also collaborate on refining the treatment of PFI. ONS are currently examining PFI contracts in some detail and will discuss the results of this with the NIESR researchers.

4.5 Data requested from DH

The team will need access to the data on expenditures and units costs underlying the ONS estimates, in particular details of the calculations to produce the HSCI. Information on prices and quantities of hospital drugs, possibly from IMS would be useful. Access to the PACT data in electronic form might also be useful; at present only a paper version for the most recent year is readily available. Data from Inland Revenue enquiries will also be needed.

4.6 R&D and Training

Since the outputs from Research and Development (R&D) and Education and Training (E&T) will not be captured in our output measure, and would prove difficult to quantify given the time scale of the study, the alternative of excluding inputs devoted to these activities will be employed. This was the strategy also used by ONS in their recent estimates of productivity growth in the NHS (ONS, 2004).

In dealing with R&D and E&T we will be guided by the research carried out by ONS. As part of its work in developing a set of Health Accounts for the UK, ONS has compiled estimates of total health expenditure in line with the framework set out in OECD (2000). In order to comply with these recommendations ONS needed to estimate spending on R&D and education and training (E&T). A description of the methodology for estimating expenditure on R&D and E&T is given in ONS (2003). Estimates of R&D expenditure were obtained from an annual government survey, the aggregate results of which are published in *Science, Engineering and Technology (SET) Statistics* (Office of Science and Technology, 2002). ONS then divided these

expenditures into labour, capital and intermediate inputs.

To account for expenditure on education and training, ONS were required to take a different approach. For England, in collaboration with the Department of Health, ONS identified items of E&T expenditure from the Department of Health and NHS budgets. An additional source of information will be information on training costs gathered by PSSRU (http://www.pssru.ac.uk/pdf/uc2003/uc2003.pdf).

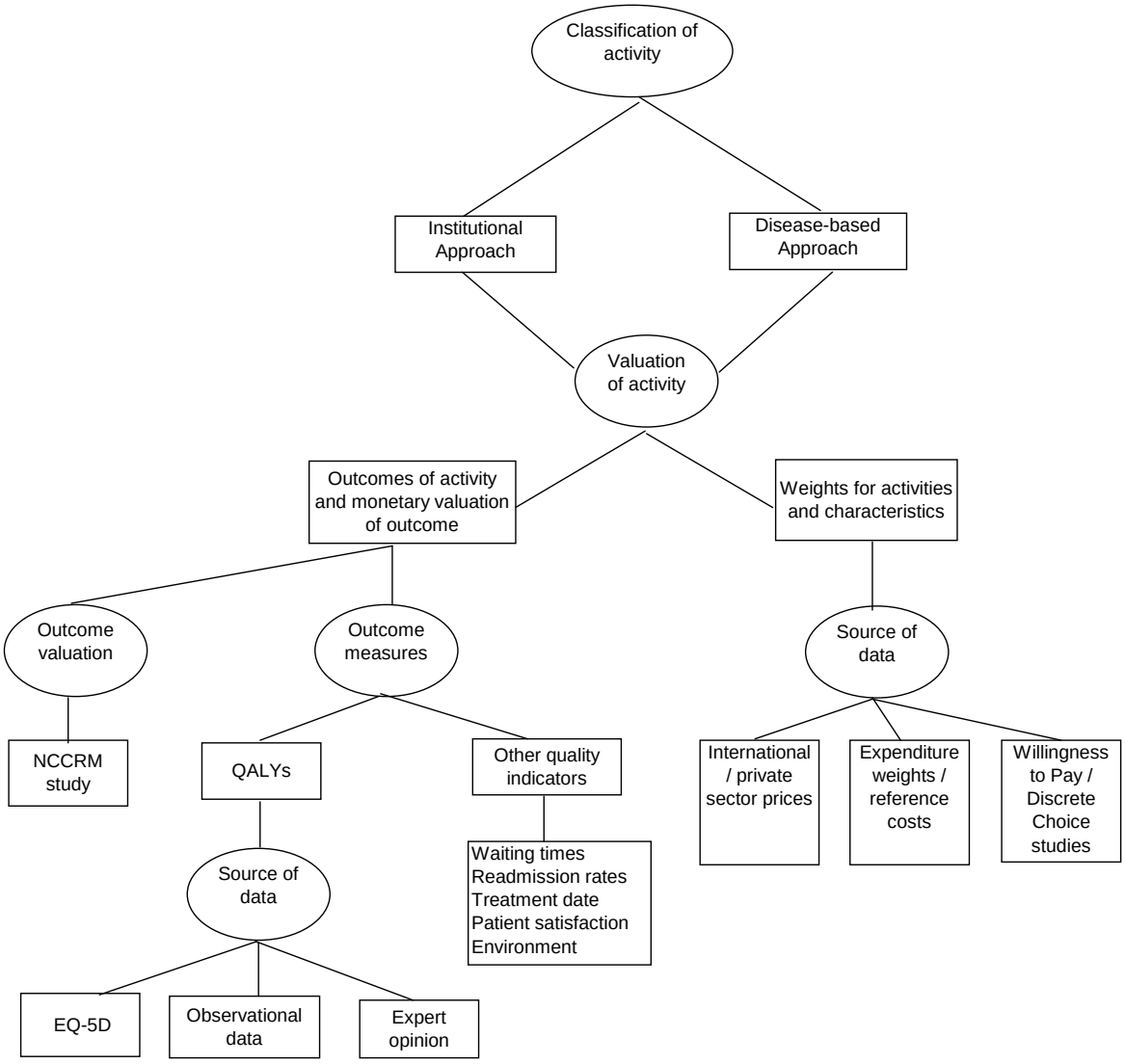
5 Conclusions and implications for phase three

Since our First Interim Report, we have focused on identifying data that can be used to measure activity in primary care, quality in secondary care, and inputs.

Figure 5.1, which is a slightly amended version of the figure in our original research proposal, maps the data required for generating a quality adjusted output index and the key decisions to be made. Given the lack of a patient identifier that links treatment of a patient across institutional settings, there is no choice but to adopt an institutional approach in the analysis of activity. However, in phase three we will look at selected disease groups to see how far it is possible to progress with UK data using the framework most commonly adopted in the health productivity literature. Given the data that we have identified thus far we expect to concentrate on orthopaedics, coronary heart disease and, possibly, arthritis.

Valuation of activity is central to a quality adjusted output measure. We identified two alternative approaches. The right hand route in Figure 5.1 looks for data to directly value changes in NHS activity. The left hand route separately quantifies the outcomes of NHS activity and then values them. With current data availability we will have to combine these two approaches.

Figure 5.1



With respect to the right hand branch of Figure 5.1, our review of available data suggests that willingness to pay studies or discrete choice experiments are too infrequent and relate to such limited activity that they cannot be used as a source of weights for an NHS wide output index intended to be updated on a regular basis. The same applies to UK private health sector prices. International prices, if published, are subject to distortions that imply that they cannot be interpreted as representing marginal social values. However, it may be possible to use existing data to examine the sensitivity of some components of the output index to the different approaches to valuation. We intend to explore this possibility further.

There are also methodological problems with using reference costs to measure the

relative value of different activities to individuals in the UK. The basic problem with using reference costs to determine the relative value of different activities is that they do not reflect quality change unless one assumes that more expensive treatments are of higher quality than less expensive treatments.

Reference costs reflect existing accounting average total costs of production. For relative reference costs to reflect the relative value of activities at the margin two key assumptions must be made. First, the way clinicians and hospital management have allocated resources between, say, treatment of arthritis patients and treatment of patients with mental health problems, must be assumed to reflect the relative marginal social benefit of these activities. Second, relative average total costs should approximate to relative marginal costs. We will continue to examine the plausibility of this second assumption in phase three by estimating marginal costs in secondary care. However, there are not enough observations on the costs of individual Trusts to enable marginal costs to be estimated for every HRG and so any measure of output growth that attempts to apply different weights for activity in different HRGs will have to continue to be based on average costs, in the short term at least. Nevertheless, our estimates of marginal cost made at a more aggregated level may indicate whether average costs are a poor approximation to marginal costs and hence whether attempts should be made to estimate marginal costs at HRG level.

An alternative approach is outlined in the left hand branch of Figure 5.1. We will seek to directly measure quality change by monitoring changes in the attributes of NHS output valued by individuals. The attributes we will be attempting to measure in phase three are: health outcomes (QALYs), waiting time, choice and certainty of date of treatment, patient satisfaction and environment.

Early in our work it became obvious that regular monitoring of quality change in the NHS requires routine collection of data on health outcomes over a representative range of NHS activity. One possibility is to rely on post-treatment mortality data and we have reviewed the data and methods available for measuring changes in quality adjusted life expectancy for NHS patients. We will use and extend these methods in phase three of the project.

However less than 3% of all NHS hospital episodes end with death of the patient and this is why it is essential for the NHS to measure the quality of health outcomes more comprehensively for patients who may be expected to benefit from treatment along dimensions other than merely mortality. We have made suggestions on how this might be done in our data requirements note to the DH (Appendix 1).

In the meantime existing outcome data might be utilised. However these data are patchy, collected on an *ad hoc* basis by individual clinicians or hospitals, and often relate to short periods of time. We have obtained three observational data sets with health measures: the Health Outcomes Data Repository (HODaR) database covering all major activity at a large Welsh NHS Trust, BUPA covering high volume elective procedures in a set of private providers, and York District Trust covering elective orthopaedic procedures. In phase three we will analyse these data to inform development of a NHS programme for routine collection of information on health outcomes. Some of the data will also be used, for some treatments, to triangulate the other methods of estimating output growth that will be applied to all secondary care activity.

For the four non-health quality relevant outcomes, there are problems with existing data that we have noted in this report. The main problem is identifying ways of measuring the relative value of these different dimensions of quality, so that estimates of changes in quality can be combined with the information on activity growth. We have found little usable price and revealed preference data, and in phase three we will concentrate on examining the sensitivity of results to different assumptions on weighting and will make further recommendations as to how the DH may proceed in future to improve the information available.

We have examined data available for the measurement of inputs. We have identified data sources for the three broad input categories, labour, intermediate and capital. In all three cases, combining a range of data sources should yield better estimates than available to now. Thus for labour input we will combine DH sources such as the NHS employment census and the NHS earnings survey with national sources such as the Labour Force Survey and the Annual Survey on earnings and hours. Intermediate and capital inputs will use data from Trust Financial Returns, Inland Revenue Inquiries,

PACT, NHS estates and national sources on historical investment series.

The measurement of inputs will proceed in collaboration with ONS. This will take the form of a series of meetings with ONS personnel where the project team will present the results from each component of the analysis. Thus it is hoped that the final estimates on the input side will be consistent with those produced by ONS.

References

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

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

Bruce, J., Russell, E.M., Mollison, J. and Krukowski, Z.H. (2001) 'The measurement and monitoring of surgical adverse events', *Health Technology Assessment*, 5 (22). Available at: http://www.hta.nhsweb.nhs.uk/fullmono/mon522.pdf

Bryan, S. and Dolan, P. (2004) 'Discrete choice experiments in health economics: for better or for worse? *The European Journal of Health Economics*, 5 (3): 199-202.

Carlson, J.A. and Parkin, M. (1975) 'Inflation expectations', *Economica*, 42: 123-138.

CDRweekly (2002) 'The first year of the Department of Health's mandatory MRSA bacteraemia surveillance scheme in acute NHS Trusts I England: April 2001 – March 2002', Health Protection Agency, 12 (25), 20 June 2002.

CDRweekly (2003) 'The second year of the Department of Health's mandatory MRSA bacteraemia surveillance scheme in acute NHS Trusts I England: April 2002 – March 2003', Health Protection Agency, 13, (25), 19 June 2003.

CDRweekly (2004) 'The third year of regional and national analysis of the Department of Health's mandatory MRSA bacteraemia surveillance scheme in acute NHS Trusts I England: April 2001 – March 2004', Health Protection Agency, 14 (29), 15 July 2004.

Commission for Health Improvement (2003) 'NHS performance ratings acute trusts, specialist trusts, ambulance trusts 2002/2003', Commission for Health Improvement, London. Available at: http://www.chi.nhs.uk/ratings/

Crowcroft, N. S. and Catchpole, M. (2002) 'Mortality from methicillin resistant Staphylococcus aureus in England and Wales: analysis of death certificates', *British Medical Journal,* 325: 1390-1391.

Dawson, D., Gravelle, H., Kind, P., O'Mahony, M., Street, A. and Weale, M. (2004a) 'Developing new approaches to measuring NHS outputs and productivity', CHE Technical Paper 31, 30 July 2004. Available at: http://www.york.ac.uk/inst/che/tech.htm

Dawson, D., Gravelle, H., Kind, P., O'Mahony, M., Street, A. and Weale, M. (2004b) 'Measurement of NHS outputs and productivity growth: memorandum to Department of Health on data requirements', 30 September 2004.

Department of Health (2001) 'NHS Performance Ratings: Acute Trusts 2000/01', Department of Health, London. Available at: http://www.doh.gov.uk/performanceratings/2001/index.html

Department of Health (2002a) 'NHS Performance Ratings and Indicators: Acute Trusts, Specialist Trusts, Ambulance Trusts, Mental Health Trusts 2001/02', Department of Health, London. Available at: http://www.doh.gov.uk/performanceratings/2002/index.html

Department of Health (2002b) 'Getting Ahead of the Curve: a strategy for combating infectious diseases (including other aspects of health protection)', Department of Health, London. Available at: http://www.dh.gov.uk/assetRoot/04/06/08/75/04060875.pdf

Department of Health (2003) 'Surveillance of Healthcare Associated Infections', Publication and Statistics, Letters and Circulars, Department of Health, London. Available at: http://www.dh.gov.uk/assetRoot/04/01/34/10/04013410.pdf

Department of Health (2004a) 'Chief Executive's Report to the NHS – Statistical Supplement May 2004', Department of Health, London. Available at: http://www.dh.gov.uk/assetRoot/04/08/26/27/04082627.pdf

Department of Health (2004b) 'Infection control training for over one million NHS staff', Publication and Statistics, Press Release, Department of Health, London. Available at: http://www.dh.gov.uk/PublicationsAndStatistics/PressReleases/PressReleasesNotices/fs/en?CONTENT_ID=4093432&chk=7bcARL

Department of Health (2004c) 'NHS Reference Costs', Department of Health, London. Available at: http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/FinanceAndPlanning/NHSReferenceCosts/fs/en

Dolan, P., Gudex, C., Kind, P., Williams, A. (1994) 'The measurement and valuation of health', Centre for Health Economics, University of York.

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

Harman, H.H. (1976) *Modern Factor Analysis*, Third Edition, University of Chicago Press, Chicago.

Healthcare Commission (2004a) '2004 performance ratings', Healthcare Commission, London. Available at: http://ratings2004.healthcarecommission.org.uk/

Healthcare Commission (2004b) 'Patient survey report 2004 - adult inpatients', Healthcare Commission, London. Available at: http://www.healthcarecommission.org.uk/assetRoot/04/00/81/82/04008182.pdf

Health Protection Agency and Office for National Statistics (2004) 'Trends in MRSA in England and Wales: analysis of morbidity and mortality data for 1993 – 2002', *Health Statistics Quarterly,* 21 – Spring 2004. Available at: http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=6725

Health Protection Agency (2003) NINSS Partnership, 'Surveillance of Surgical Site Infection in English hospitals 1997 – 2002'. Available at: http://www.hpa.org.uk/infections/topics_az/hai/SSIreport.pdf

Health Protection Agency (2004) 'Protocol for Surveillance of Surgical Site Infection', Surgical Site Infection Surveillance Service, England. Available at: http://www.hpa.org.uk/infections/topics_az/hai/SSI%20Protocol.pdf

Hippisley-Cox, J. (2004) 'Description of the National QRESEARCH Database' (4th Download), Report to the Statistics Division, Department of Health, Report 14, version 1.4, 8 November.

Hofer, T.P. and Hayward, R.A. (1995) 'Can early re-admission rates accurately detect poor-quality hospitals?', *Medical Care,* 33 (3): 234-45.

Hurst, J. and Siciliani, L. (2003) 'Tackling Excessive Waiting Times for Elective Surgery: A Comparison of Policies in Twelve OECD Countries'*, OECD Health Working Paper 6*, OECD, Paris. Available at: www.oecd.org/health

Ludke, R.L., Booth, B.M. and Lewis-Beck, J.A. (1993) 'Relationship between early readmission and hospital quality of care indicators', *Inquiry,* 30: 95-103.

Mai, N. (2004) 'Measuring health care output in the UK: a diagnosis based approach', Office for National Statistics.

Mitchell, J. (2002) 'The use of non-normal distributions in quantifying qualitative survey data', *Economics Letters*, 76: 101-107.

Munro, J., Nicholl, J., O'Cathain, A., Knowles, E. (2000) 'Evaluation of NHS Direct first wave sites: second interim report to the Department of Health'. Medical Care Research Unit, University of Sheffield, March 2000.

Murray, C.J.L. and Lopez, A.D. (1996) *The Global Burden of Disease and Injury*, Harvard University Press, Boston.

OECD (2000) *A System of Health Accounts*, OECD, Paris. Available at: http://www1.oecd.org/publications/e-book/8100061e.pdf

OECD (2001) *Measuring Productivity: Measurement of Aggregate and Industry Level Productivity Growth*, OECD, Paris. Available at: http://www.oecd.org/dataoecd/59/29/2352458.pdf

Office of Science and Technology (2002) 'Science, Engineering and Technology (SET) Statistics.' Available at: www.dti.gov.uk/ost/setstats/figtab.htm

O'Mahony, M. and van Ark, B. (2003) *EU Productivity and Competitiveness: An industry perspective. Can Europe resume the catching-up process?* European Commission, Enterprise publications.

ONS (2003) 'Total UK health expenditure', Health Expenditure methods, December 2003, National Statistics. Available at: http://www.statistics.gov.uk/healthaccounts/downloads/HealthexpendituremethodsDec2003.pdf

ONS (2004) 'Public Service Productivity: Health', National Statistics, October. Available at: http://www.statistics.gov.uk/articles/nojournal/public_service_productivity_health.pdf

Roland, M. (2004) 'Linking physicians' pay to the quality of care: a major experiment in the United Kingdom', *New England Journal of Medicine*, 351:1448-1454

Siciliani, L. and Hurst, J. (2004) 'Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries', *Health Policy* (forthcoming).

Stouthard, M.E.A., Essink-Bot, M.L., Bonsel, G.J., Barendregt, J.J., Kramer, P.G.N., van de Water, H.P.A., Gunning-Scheppers, L.J. and Van der Maas, P.J. (1997) 'Disability weights for diseases in the Netherlands', Department of Public Health, Erasmus University, Rotterdam.

Thomas, J.W. (1996) 'Does risk-adjusted readmission rate provide valid information on hospital quality?' *Inquiry,* 28: 258-70.

Triplett, J. (2004) 'Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products', STI Working Paper, 2004/9, OECD, Paris. Available at: http://www.oecd.org/dataoecd/37/31/33789552.pdf

Vallance-Owen, A. (2004) 'Routine use of health related quality of life measure (SF-36 and VF-14)', presentation to conference on Managing Consultant Performance, 9 November 2004.

Williams, A. (1985) 'Economics of coronary artery bypass grafting', *British Medical Journal*, 291: 326–329.

World Bank (1993) 'World Development Report 1993: Investing in health – world development indicators', Oxford University Press.

Acknowledgements

The outputs and productivity project is funded by the Department of Health. The views expressed are not necessarily those of the DH. We are grateful to Mark Dusheiko for assistance with HES data, to Geoff Hardman for analysing it, to Andrew Vallance-Owen and Justin Whatling for providing the BUPA health change data, and to Phillip Lee for discussions on data and methods in measuring inputs. We have also benefited from advice from David Cutler, a member of our steering group.

Appendices

Appendix 1: Memorandum on data requirements

Measurement of NHS Outputs and Productivity Growth

Memorandum to Department of Health on data requirements

30 September 2004

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

1 Introduction

At the meeting of the Steering Group for the Productivity project on 7th July 2004 to discuss a draft of First Interim Report[5] on methodology the research team were asked to produce a memorandum on data requirements in advance of its second interim report on data timetabled for 30th November 2004.

This memorandum sets out

- *requests* for data held by the DH which we know or believe exist and which we require for the investigation of improved productivity and output measures in the third phase of our project starting on 1st December 2004.
- *suggestions* for data which are currently not collected by the DH. It may not be possible for the DH to produce these data in the timescale of our project but we believe they are essential for measurement of productivity and output and that they would also be extremely useful for other purpose such as monitoring performance of lower level units in the NHS.

We appreciate that the Department of Health is trying to reduce the amount of data

[5] Dawson D, Gravelle H, Kind P, O'Mahony M, Street A, Weale M. Developing new approaches to measuring NHS outputs and productivity. CHE Technical Paper 31, 30 July 2004. http://www.york.ac.uk/inst/che/tech.htm

collected from the NHS (Review of Central Returns Unit). However, the data requirements outlined below are essential to development of robust measures of output, productivity and quality change in the NHS. These data will be required not only by the DH but also by outside bodies such as the Treasury and ONS.

The memorandum does not cover

- the data set used by EOR in the construction of the revised NHS output index. We have had ready access to some of these data and assume that will be no problems with obtaining further information as required.
- data that we already have from the DH or from other sources (eg HES data).

2 Outcomes Data

2.1 Health outcomes

The main aim of the health system is the improvement of the health of the population. This being so, it would seem reasonable that any measure of health system productivity should include measures of the effect of the system on health. The challenges associated with measuring the effect of interventions are detailed and discussed in our first interim report.

The construction of a productivity index requires information about changes in health status attributable to interventions. Such information currently is not collected by the NHS. We *suggest* the systematic use of a standardised measure of health status to improve the effective management of the NHS and to provide the fundamental data needed to properly reflect changes in NHS productivity.

The following points need to be addressed in such a data collection exercise

- **NHS patient sample**. Since the scale of NHS activity is so broad and the potential volume of patients is so large sampling seems a more sensible strategy than attempting to measure health effects for all NHS patients in a sector. Sample sizes are likely to vary across different types of patients. While

a random sample of the NHS patient population would be preferable, in the first instance it may be satisfactory to undertake a pilot exercise at a handful of Trusts. Even in the longer term there may well be economies of scale from concentrating data collection within a small number of institutions rather than spreading the data collection burden more thinly across all Trusts. The drawback is that the selected Trusts may not be fully representative of the NHS as a whole.

- **Choice of instrument**. A single generic, rather than condition-specific instrument is required in order to facilitate aggregation across different types of NHS activity. Profiles such as SF-36 provide multiple measures of outcome but are unsuitable for most non-clinical purposes since they typically lack the capacity to form a single aggregate index. Derivatives of SF-36 such as the SF-6D do not suffer from this deficiency. The EQ-5D is designed to produce a single index and its five dimensions have been calibrated in terms of social preference weights of a UK population and is probably the primary candidate measure.
- **Timing**. The timing of before and after health status measurement may depend on the type of activity (emergency or elective). Measurements from patient self-report are preferable but there are circumstances in which proxy or retrospective data provide the only feasible route (for example, when emergency care is required by patients in an unconscious state). The timing of the administration of the instrument may also depend on diagnostic category or intervention type since different treatments may have an effect over shorter or longer periods. For some activities or conditions it may be necessary to make several post interventions measurements since the effect on health may vary substantially with the time elapsed since treatment.
- **Grouping of NHS activities**. Given the enormous range of NHS activities it is necessary to group them for data analysis. Thus the main grouping of secondary care activities is by HRG which attempts to group activities by their costs. But a given HRG may contain a large number of procedures which have very different effects on health. The availability of patient-level health outcomes data matched to other datasets (such as HES) will make it possible to explore the extent to which health outcomes are related to other routinely collected patient characteristics, such as age, gender, diagnoses and

procedures. This information may be used to create homogenous health outcome grouping which, in turn, may allow refinement of how activities are defined.

- **Frequency of data collection.** If the effect of treatment on health depended only on the state of medical knowledge and the pace of technological change in medicine was slow enough it could be argued that collecting data on health effects was an exercise that needed to be undertaken at intervals of several years. But technological change in medicine and pharmaceuticals is rapid, the NHS is subject to frequent organisational change which may affect the mix of patients receiving particular treatments and the speed at which new technology spreads. We believe that only a continuous sampling of the NHS patient population will be adequate to capture trends in the effects of the NHS on its patients.

2.1.1 Feasibility

Outside clinical trials, experience of routine collection of health status data in the UK is patchy. Individual clinicians and clinical teams make use of a variety of standardised measures, but this is largely uncoordinated, its coverage remains undocumented[6] and aggregation of such data is problematic given the use of different instruments. The feasibility of collecting more useful data on health status has two components – identifying an appropriate means of measuring health status and designing a mechanism for capturing those measurements.

Although there are a limited number of examples of prospective health data collection these examples demonstrate that such data collection would be feasible in the NHS.

- The survey of acute in-patients conducted by **Picker International** showed that it was possible to collect EQ-5D data from a sample of patients recently discharged from all NHS Trust hospitals. The value of these data would have been enhanced if they been linked with basic HES variables, such as diagnosis or healthcare resource group.

[6] A catalogue of NHS-based users of EQ-5D is in preparation

- The **Health Outcomes Data Repository (HODaR)** operates a continuous survey of all in-patients and out-patients at a single large Welsh Trust. These are now linked to individual level primary and community care data. Data for more than 30,000 patients have been collected. Almost 10% of these having completed EQ-5D on more than one occasion. However, the data are predominantly based on post-discharge observations and this limits their value in measuring health outcomes. Since the advent of this current project the HODaR survey has started to collect data on pre-admission health status.

- For a number of years **BUPA** has been routinely administering health status questionnaires to patients before and three months after treatment, with some 100,000 having now been surveyed. These data are restricted primarily to elective procedures. BUPA plan to extend them to four types of of cancer.

The HODaR data set includes SF-36 and EQ-5D. BUPA collects data mainly on SF-36 with some data on a visual acuity scale for cataract patients. We have access to the HODar EQ-5D data and we are currently negotiating with BUPA for similar access to their datasets. The data will be useful in order to derive estimates of the sample sizes required for future collection of this type of data by the NHS.

Further evidence on feasibility should emerge from the PROMS studies being undertaken by the London School of Hygiene and Tropical Medicine. These studies will provide information on a limited set of surgical procedures conducted in treatment centres. It would be important to undertake similar exercises for medical conditions and for activities conducted in other settings.

2.1.2 Cost

The incremental costs of introducing systematic observation of health status via existing information systems is difficult to estimate. It would seem sensible to consider an extension to the current HES-based data to provide maximum scope for exploitation through record linkage. Modification of this sort ought not to incur a significant cost. However, the data capture from patients will require additional

organisational and administrative costs. Patient-centred reporting systems using traditional paper and pencil techniques require costly processing in order to link them to other NHS data. Computer-assisted interview methods have scope for more efficient data acquisition and transmission but would need more costly administration. The use of handheld PDA recording systems is now becoming a feature of many clinical trials that record patient-reported health status and it can be expected that hardware costs will continue to fall.

Currently BUPA estimates that it costs around £4 per patient to administer their manual system of health status measurement based on SF-36.

The efficiency of current patient-survey questionnaires could be improved without cost if a standard measure of health status formed part of all schedules. The present survey of patients in primary care contains a form of SF-36 whilst in-patient surveys incorporate EQ-5D.

The introduction of systematic health status measurement might be achieved under the aegis of the National Programme for Information Technology announced in December 2002 with a budget of £2.3 billion and which the Audit Commission suggested would provide the Department of Health with the opportunity to improve NHS data quality.

2.2 Other outcome measures: hospital mortality

Both the research team and EOR are investigating ways in which hospital mortality data can be incorporated into improved measures of outcomes in the secondary care sector. Whilst necessarily cruder than health status measures, mortality data are an important aspect of outcome measurement. In the short to medium term, before systems are introduced to collect health status measures, mortality is the only measure of the health effects of treatment.

We have, via HES, data on patients who die in hospital but it would be better to have information on deaths within a set period of discharge as well. It is planned that such

information will be available in HES but it is not yet available. EOR have provided us with data on total deaths within 30 days by age and sex. We would like to have the data on deaths within 30 days disaggregated to diagnosis (ICD code) or HRG.

2.3 Other outcome measures: patient satisfaction

As we emphasised in our First Interim Report the effect of the NHS on health status is obviously an extremely important dimension of NHS outcomes but other dimensions, especially process related outcomes, should also be taken into account. The NHS has undertaken a series of national surveys of patients (NHS Patients Survey Programme 1998 and 2002 which covers general practice, and the Acute Inpatient Survey 2001/2) which elicit patient views on the NHS experiences. We would like to have access to the individual level data in these surveys in order investigate whether it is possible to use patient views on various dimensions of care construct measures of the quality of care.

3 Primary Care Data

3.1 General practice

Currently two measures of activity in general practice are included in the revised NHS outputs index: consultations and prescriptions. In previous years the CWAI did not include consultations and prescriptions but did include such activities as cervical screening tests, and visits by district nurses and health visitors.

3.1.1 Consultations

Estimates of consultation activity are derived from the consultations reported by respondents in the General Household Survey and are available by location (surgery, home, phone) and provider (GP, practice nurse (but only after 2000)). The estimate of the number of consultations per year is made by multiplying the number of reported consultations in the 14 days prior to interview by 26. No allowance is been made for seasonal factors - the date of the consultation varies across respondents and has also

varied between rounds of the GHS. Estimates of the number of consultations based on the GHS are around 20% below those from studies based on GP record systems. There have also been implausibly large changes in the numbers of consultations reported in the GHS for some age-gender groups from one year to next. For these reasons it is preferable to base estimates of GP activity on GP record systems.

Ideally NHS productivity measures should be based on numbers of patient journeys of different types where journeys are likely to involve both primary and secondary care. In the absence of routine record linkage such measures are not currently feasible but it would still be worthwhile getting a finer breakdown of GP consultations to allow for the changing mix of providers and for the changing mix of types of consultations.

We *suggest* that the DH embark on a programme to extract data from GP record systems to give the number of consultations in a year broken down by

- provider (GP, nurse – any other providers in the practice – eg physiotherapy)
- age, gender of patients to enable the data to be grossed up to yield national estimates. (Ideally we would want to have richer data on patients on characteristics known to affect consultations which are also measured in the 2001 Census such as ethnicity, education but doubt if such data is recorded reliably in GP record systems.)
- diagnostic category
- result of consultation – prescription, referral, monitoring of condition, advice etc

In order to gross up from the sample of records to yield national estimates we would want, in addition to the data on age, gender etc of patients, information on the numbers and types of practice (location, socio-economic characteristics of area, size of practice, GP numbers, GP age and gender). The richer the socio-economic data on consulting patients the less critical is data on practice population characteristics.

In order to determine if this level of detail on consultations makes any difference to estimates of productivity growth we would like data for say 5 years. It may be that some databases which have a reasonable run of historical data have poorer data than other sources which have been more recently established so that we might want to

have data from more than one data base.

Data availability. There are a number of databases deriving information from GP record systems which may be able to provide the information required: GPRD, QRESEARCH, RCGP Weekly Returns, IMS, Practice Team Information (previously Continuous Morbidity Recording from 60 Scottish practices. The research team and EOR are in the process of consulting with experts in PRIMIS and elsewhere with detailed knowledge of the data bases and record systems to investigate how suitable these sources are and what useful data could be collected from GP record systems on GP activity.

3.1.2 GP Cost weights

The PSSRU estimates the unit costs of GP and nurse consultations (http://www.pssru.ac.uk/pdf/uc2003/uc2003.pdf) using from a variety of official and unofficial sources. Several of the estimates rest on self reported GP activity from the 1992/3 GP Workload Survey undertaken for the DDRB. There does not appear to be a more recent survey of GP activity and we *suggest* that DH should consider undertaking such a survey are regular intervals.

3.1.3 Prescribing

The prescription activity measure in the recently revised NHS outputs index is derived from PPA data. The PPA data are collected in order to remunerate pharmacists (and dispensing GPs). It is therefore a comprehensive measure of prescriptions dispensed and can be disaggregated to product type if required. The data are reliable, comprehensive and readily available at national levels of aggregation. They have been used to construct a number of indicators of practice prescribing quality as well as quantity.

The usefulness of the data could be greatly improved and this would be relatively simple. The most obvious example is by improving the patient information on the prescription form. At the moment the only patient data on the form indicates if the patient is entitled to free prescriptions and on what grounds. The information has

been used by the Prescribing Support Unit to produce the Low Income Scheme Index which measures the proportion of prescriptions which are dispensed without charge on grounds of low income. The LISI is the only direct variable measuring practice population socioeconomic status which relates directly to practice patients rather than being attributed from Census or Social Security data on the basis of patient postcode. Adding a field for diagnosis to the prescription form would greatly enhance the usefulness of routine prescribing data as a measure of prescribing quality. Adding gender and age fields would also improve the socioeconomic data and improve prescribing quality indicators. We suggest that the DH should consider adding these fields to the prescription form.

3.1.4 Other GP activity data

The data to be collected by QMAS/QPID based on the QOF introduced in the new GMS contract will enrich the set of activities which are routinely measured. However, the data will be collected for the first time in 2004/5 and so cannot be used in calculations of productivity growth until the 2005/6 data are available, which will be after the end of the current project. Estimates of productivity growth based on QMAS data are likely to be unreliable for the first few years of the new contract because of changes in data collection and coverage. It is also unclear how good the QOF coverage of the 35% of practices with PMS contracts will be. The data on targeted activities and services rewarded under the old and new contract is available, though there appear to be problems with reliability of the data from practices with PMS contracts. We *suggest* that the DH should attempt to ensure that the QOF and targets data from PMS practices is comparable in quality and coverage to that for nGMS practices.

3.2 Other primary care data

NHS Direct, NHS Direct Online and Walk-In Centres are recent innovations in the provision of first contact advice and information. They are likely to reduce the costs to patients of such first contacts, leading both to an increase in primary care activities and to a change in the mix of activities in general practice. The organisations are

expected to play an increasing role in the NHS over the coming years and it is important that their presence is recognised in measures of NHS output and productivity.

3.2.1 NHS Direct and NHS Direct Online

Aggregate data on use of NHS Direct and NHS Direct Online are available, as reported in the Chief Executive's Report to the NHS – Statistical Supplement (May 2004). The data differ slightly from those in the spreadsheets compiled (29 March 2004) by EOR for the revised NHS output index.

In order to measure the outputs of the services more accurately it would be helpful to have data on

- the breakdown of enquires between the provision of health advice and information about the health service
- type of conditions people seek health advice about
- actions are recommended as a result of the request

It is possible that such data have been collected, for example via the website service for those who seek advice from a nurse which involves self-completion of a detailed questionnaire on the nature of the symptoms and condition, as well as personal information. Presumably – although we have not been able to ascertain whether this is the case – enquiries that result in a self-care recommendation are logged also. The telephone service seems to be set-up in a similar fashion, the difference being that the information is recorded by NHS Direct staff.

It appears, therefore, that these two organisations routinely collect (or, at least, have the capacity to collect) detailed electronic information from every person making an enquiry about their (or their family member's) health condition.

We have been unable to obtain such information from NHS Direct or to find out precisely what is available.

3.2.2 Walk-in Centres

We have data on total visits from the spreadsheets compiled (29 March 2004) by EOR for the revised NHS output index. They differ slightly from those in the Chief Executive's Report to the DH, 2004, Statistical Supplement (table 2.2.2).

We are awaiting financial data and more detailed activity data.

4 Inputs

It is important to have a comprehensive coverage of inputs used in producing health services in order to explain changes in outputs and to measure productivity. Thus we require both volume measures (e.g. numbers of doctors) and prices of inputs (e.g. wage payments to doctors) and measures of the extent to which input qualities are changing. Under the assumption that wages equal marginal products, disaggregation of input groups by type can yield some information on quality change. Where these assumptions are not met, additional information may be required.

For example the employment of whole time equivalent (wte) hospital doctors increased on average by 3.3 per cent per annum from 1995 to 2003. In the same period the employment of wte consultants rose by 4.3% p.a. Since consultants earn significantly more than other doctors, and by implication add more to output, treating each doctor as equivalent will understate growth in 'doctor input'. A quality adjustment to the growth in total doctors can be derived by dividing doctors by type and then weighting each type by their shares in total doctor wage payments. A rough calculation, based on wage shares of consultants and all other doctors, gives a 'quality adjusted' growth for doctors equal to 3.6% per annum, an upward adjustment of nearly 10%.

It is common practice to divide inputs into labour, capital and intermediate inputs. This section considers each in turn, setting out data required, known data sources and

gaps that require additional information. It then discusses other data requirements to take account of training and R&D activities.

4.1 Labour

4.1.1 Requirements

Number of wte workers and wage payments by type (doctors, nurses, AHPs, HCAs and other).

4.1.2 Available data

Comprehensive data on wte numbers from NHS workforce census; additional information on numbers cross-classified by qualifications from the Labour Force Survey (LFS).

Information on wage and salary payments for types of clinical staff and other staff are available from Financial Returns (TFR, PFR and HFR). Data on wages by type and qualification are also available from the LFS and the New Earnings Survey.

4.1.3 Gaps in data requirements

It should be possible to link these data to derive a reasonable measure of labour input. Nevertheless a better measure could be derived if wage information was more tightly linked to the employment data in the NHS workforce census. The published data from financial returns are less detailed than the NHS workforce census. For example the TFRs contain sufficient detail by type for hospital doctors but not other staff. In particular nurses are divided into only two categories, senior nurse and other. It would be useful to have wage and salary payments by detailed type corresponding to the categories in the NHS workforce census. We understand that there is an annual earnings survey for the NHS[7] that may contain such data but so far we have not been

[7] ROCR ongoing approved non-finance data collections 2004-05 as at July 2004: http://www.dh.gov.uk/PublicationsAndStatistics/Statistics/StatisticalCollection/StatisticalCollectionArticle/fs/en?CONTENT_ID=4031147&chk=3ff07c

able to access these data.

We also require information on GP income and expenses (practice expenses, etc). We understand theses data are collected by Inland Revenue for a sample of GPs at the request of the DDRB. The DDRB no longer prices the GMS contract. Does this mean that it will no longer acquire Inland Revenue data on GP income and expenses? Does the Department of Health have other arrangements in place to collect this data from Inland Revenue?

4.2 Capital

4.2.1 Requirements

Annual data on nominal investment, distinguishing the important asset types employed in the NHS and asset deflators.

4.2.2 Available data

Annual investment data, distinguishing buildings, plant & machinery, transport equipment and information technology are available from NHS estates. Age profile and quality indicators for buildings can also be obtained from this source. The TFRs contain information on expenditure on medical and surgical equipment, X-ray equipment, laboratory equipment and appliances in hospitals from TFRs. Deflators can be obtained from ONS, but these will not generally be specific to the NHS.

4.2.3 Gaps in data requirements

In order to quality adjust capital input it would be useful to have separate investment data on types of equipment that have seen rapid technological change and change in unit cost (e.g. MRI scanners). Alternatively it would be useful to have the value of the stock of these assets, numbers of items and age profiles of the stock. One problem that must be addressed is the gap in data created by PFI confidentiality. We understand that a significant proportion of new investment in equipment such as scanners is being undertaken under PFI contracts. Unless it is possible to access information on stocks and value, it may not be possible to adequately deal with questions of productivity

growth and technical change associated with investment in new equipment. We also require investment by GPs.

4.3 Intermediate inputs

4.3.1 Requirements

Expenditure and deflators by main categories, i.e. drugs, clinical supplies, other supplies, fuel and power, purchased services.

4.3.2 Available data

Financial returns cover nominal expenditures for the broad categories outlined above. Information on prescription drugs, volumes and unit costs, are available from PACT. Data on hospital drugs (volumes and list prices) may be available from IMS health. As non-subscribers, we cannot access this information. Does the DH have access to the data? Deflators for other categories of intermediate input can be obtained from ONS but these are unlikely to be NHS specific.

4.3.3 Gaps in data requirements

The NHS obtains hospital drugs at significant discounts so that list prices are unlikely to be very useful. The complexities of the pharmaceutical market are such that the assumption that price reflects quality is unlikely to hold. Therefore additional information is required. One possibility is to examine patient outcomes and drug use from disease registers or to obtain opinions from panels of experts. This is simply another example of the need for outcomes data if we are to measure technical change and productivity in health care.

4.4 Training & R&D

Ideally we would like to measure training and R&D outputs but this is unlikely to be feasible given the time scale of the report. An alternative is to adjust inputs to take out those used in these two activities.

4.4.1 Requirements

We require estimates of the time spent by NHS staff in training, net of benefits provided by trainees, plus any other inputs used specifically for training; costs and staff time, capital and intermediate input use in R&D activities.

4.4.2 Data Availability

We understand there is an internal DH report that attempts to quantify service and training costs for hospital doctors. This may be of use in our analysis of productivity change if it could be made available. It may also be possible to make some use of the information on training costs gathered by PSSRU (http://www.pssru.ac.uk/pdf/uc2003/uc2003.pdf).

4.4.3 Gaps in data requirements

Comprehensive assessments of training staff time and costs, and information on R&D costs.

Appendix 2: Community care data

A1 - Ambulance

Data content	Years					
	2003-04	2002-03	2001-02	2000-01	1999-00	1998-99
Summary statistics of emergency calls to ambulance services and patient journeys, England	*	*	-	-	-	-
Number of patient journeys by priority of journey, time series from 1988-89 to 2001-02	-	-	*	*	*	*
Number of patient journeys by priority of journey and Ambulance Service	*	*	*	*	*	*
Total Number of Emergency calls, time series from 1994-95 to 2003-04	*	*	*	*	*	*
Emergency calls resulting in emergency response arriving at the scene of the incident by Ambulance Service, time series from 1994-95 to 2003-04	*	*	*	*	*	*
Emergency calls: response times, by Ambulance Service	*	*	*	-	-	-
Emergency calls: response times for services without call prioritisation, by Ambulance Service	-	-	-	*	*	*
Emergency calls: response times for services with call prioritisation, by Ambulance Service	-	-	-	*	*	*
Emergency calls: responses within 14/19 minutes, by category of call and Ambulance Service, time series from 1994-95 to 2003-04	*	*	*	*	*	*
Emergency calls: responses within 8 minutes, by category of call and Ambulance Service, time series from 1994-95 to 2003-04	*	*	*	*	*	*
Arrival times for urgent journeys, by Ambulance Service	*	*	*	*	*	*
Arrival times for urgent journeys, by Ambulance Service, time series from 1998-99 to 2003-04	*	*	*	*	*	*
Urgent Standard Response times: arrival time in relation to requested arrival time: arrival not more than 15 minutes late, by Ambulance Service, time series from 1994-95 to 2002-03 1996-97 to 2001-02	-	*	*	*	*	*
Urgent journeys, by Ambulance Service, time series from 1994-95 to 2003-04	*	*	*	*	*	*

* = available

- = not available

A2 – Chiropody

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, by region, time series 1988-89 to 2000-01	-	-	*
Contacts, time series from 1988-89 to 2002-03 (combined data)	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
Initial contacts (new episodes of care) by age	*	*	-
Initial and first contacts by Provider	*	*	*

* = available

- = not available

A3 - Clinical Psychology Services

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, by region, time series 1988-89 to 2000-01	-	-	*
Contacts, time series from 1988-89 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral	*	*	*
Initial contacts (new episodes of care) by source of referral and region	-	-	*
Initial contacts (new episodes of care) by source of referral, time series form 1988-89 to 2000-01	-	-	*
Initial contacts (new episodes of care) by age and sex	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
Initial and first contacts by Provider	*	*	*

* = available

- = not available

A4 - Community Mental Health Nursing

Data content	Years		
	2002-03	2001-02	2000-01
Initial and First contacts by provider, 2000 –01	-	-	*
Contacts, time series form 1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral and region	-	-	*
Initial contacts (new episodes of care) by source of referral, time series from 1988-89 to 2002-03 (combined data)	*	*	*
Initial contacts (new episodes of care) by age and sex	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
First contacts (different persons receiving care) by age and sex	*	*	*
First contacts (different persons receiving care) by age and region	-	-	*
First contacts (different persons receiving care) by age, time series from 1988-89 to 2002-03 (combined data)	*	*	*
Total face to face contacts by location, 1988-89 to 1999-2000	-	-	*
Initial and First contacts by provider	*	*	*

* = available

- = not available

A5 - Contraceptive Services

Data content	Years				
	2002-03	2001-02	2000-01	1999-00	1998-99
Summary statistics on contraception, time series from 1988 to 2002 (combined data)	*	*	*	*	*
First contacts at NHS family planning clinics by sex and (women only) age, 1975 to 2002-03 (combined data)	*	*	*	*	*
First contacts with women and men at family planning clinics by primary method of birth control, time series from 1975 to 2001-02 (combined data)	-	*	*	*	*
First contacts with women and men at NHS family planning clinics by primary method of birth control, time series from 1992-93 to 2002-03	*	-	-	-	-
First contacts with women at family planning clinics by primary method of birth control and age	-	*	*	*	*
First contacts with women at NHS family planning clinics by primary method of birth control and age	*	-	-	-	-
Occasions on which post-coital contraceptives dispensed at family planning clinics by type, time series from 1989-90 to 2001-02 (combined data)	-	*	*	*	*
Occasions on which post-coital contraceptives dispensed at NHS family planning clinics by type, 1992-93 to 2002-03	*	-	-	-	-
Occasions on which post-coital contraceptives dispensed at family planning clinics by type and age	-	*	*	*	*
Occasions on which post-coital contraceptives dispensed at NHS family planning clinics by type and age	*	-	-	-	-
First contacts with men at NHS family planning clinics, time series from 1975 to 2002-03 (combined data)	*	*	*	*	*
Total contacts at family planning clinics, first contacts and domiciliary visits, time series from 1975 to 2000-02.	-	*	*	*	*
Total contacts at NHS family planning clinics, first contacts and domiciliary visits, time series from 1975 to 2000-02.	*	-	-	-	-

* = available

- = not available

A5 - Contraceptive Services continued

Data content	Years				
	2002-03	2001-02	2000-01	1999-00	1998-99
Selected information on family planning clinic services by region	-	-	*	*	*
Clinic sessions for people aged under 20 (1) 1992-93 to 2002-03	*	-	-	-	-
GP prescriptions for contraceptives, 1991 to 2000	-	-	*	-	-
Contraceptive prescription items dispensed in the community, time series from 1991 to 2002	*	*	-	-	-
Sterilisations and vasectomies by age	*	*	*	-	-
Contacts by provider	*	*	-	*	*
Initial contacts (new episodes of care) by provider	-	-	*	-	-

* = available

- = not available

A6 - District Nursing

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, by region, time series 1988-89 to 2000-01	-	-	*
Contacts, time series from 1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral	*	*	*
Initial contacts (new episodes of care) by source of referral and region	-	-	*
First contacts (new episodes of care) by age and sex	*	*	*
First contacts (new episodes of care) by age and region	-	-	*
First contacts (different persons receiving care) by age, time series 1988-89 to 2002-03 (combined data)	*	*	*
Total face to face contacts by locations, time series from 1988-89 to 1999-2000	-	-	*
Initial and first contacts by Provider	*	*	*

* = available

- = not available

A7 - Health and other Professional Advice and Support Programmes in the Community

Data content	Years		
	2002-03	2001-02	2000-01
Contacts with Health Visitors by region, 1988-89 to 2000-01	-	-	*
Contacts with Health Visitors, time series from 1991-92 to 2002-03 (combined data)	*	*	-
First contacts (different persons receiving care) with Health Visitors by age and region	-	*	*
First contacts (different persons receiving care) with Health Visitors by age	*	-	-
First contacts (different persons receiving care) with Health Visitors by age, time series from 1988-89 to 2002-03 (combined data)	*	-	*
Contacts with Health Visitors by provider 2000-01	-	-	*

* = available

- = not available

A8 - Learning Disability Nursing

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, by region, time series 1988-89 to 2000-01	-	-	*
Contacts, time series from 1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral	*	*	*
Initial contacts (new episodes of care) by source of referral and region	-	-	*
Initial contacts (new episodes of care) by age and sex	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
First contacts (different persons receiving care) by age and sex	*	*	*
First contacts (different persons receiving care) by age and region	-	-	*
First contacts (different persons receiving care) by age, time series 1988-89 to 2002-03 (combined data)	*	*	*
Total face to face contacts by locations, time series from 1988-89 to 1999-2000	-	-	*
Initial and first contacts by Provider	*	*	*

* = available

- = not available

A9 - Occupational Therapy Services

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, by region, time series 1988-89 to 2000-01	-	-	*
Contacts, time series from 1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral	*	*	*
Initial contacts (new episodes of care) by source of referral and region	-	-	*
Initial contacts (new episodes of care) by source of referral, time series 1988-89 to 2002-03 (combined data)	*	*	*
Initial contacts (new episodes of care) by age and sex	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
Total face to face contacts by locations, time series from 1988-89 to 1999-2000	-	-	*
Contacts by Provider	*	*	*

* = available

- = not available

A10 - Physiotherapy Services

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, time series from 1988-89 to 2002-03	*	-	-
Initial contacts (new episodes of care) by region	-	-	*
Initial contacts (new episodes of care) by source of referral	*	*	*
Initial contacts (new episodes of care) by source of referral and region	-	-	*
Initial contacts (new episodes of care) by source of referral, time series 1988-89 to 2002-03 (combined data)	*	*	*
Initial contacts (new episodes of care) by age and sex	*	*	*
Initial contacts (new episodes of care) by age and region	-	-	*
Initial contacts (new episodes of care) by Provider	-	*	*
Initial contacts by Provider	*	-	-

* = available

- = not available

A11 - Specialist Care Nursing

Data content	Years		
	2002-03	2001-02	2000-01
Initial contacts with specialist care nurses by staff group, time series from 1994-95 to 2002-03 (combined data)	*	-	*
Initial contacts with cancer and terminal care nurses by staff group and Region	-	-	*
Initial contacts with cancer and terminal care nurses by staff group and Provider	*	-	*
Initial contacts with other specialist care nurses by staff group and Region	-	-	*
Initial contacts with other specialist care nurses by staff group and Provider	*	-	-

* = available

- = not available

A13 - Speech and Language Therapy Services

Data content	Years		
	2002-03	2001-02	2000-01
Contacts, time series from 1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral, time series from1991-92 to 2002-03 (combined data)	*	*	-
Initial contacts (new episodes of care) by source of referral	*	*	-
Initial contacts (new episodes of care) by age and sex	*	*	-
Contacts by provider	-	*	-
Initial contacts (new episodes of care) by provider	*	-	-

* = available

- = not available

Appendix 3: Screening and diagnosis data

A1 - Breast Screening Programme

Data content	Years					
	2002-03	2001-02	2000-01	1999-00	1998-99	1997-98
Summary statistics on breast cancer and the NHS breast screening programme	*	*	-	-	-	-
Test status of women and coverage by age, at 31 March	*	*	*	*	*	*
Test status and coverage of target age group 50-64 by region, at 31 March	-	*	*	*	*	*
Test status and coverage of target age group 53-64 by region, at 31 March	*	-	-	-	-	-
Test status and coverage of target age group 53-64 by Strategic Health Authority, at 31 March	*	*	-	-	-	-
Estimated uptake of invitations to screen by women aged 50-64 by region	*	*	*	*	*	*
Estimated uptake of invitations to screen by women aged 50-64 by type of invitation and region	*	*	*	*	*	*
Number of women invited and estimated uptake of invitations to screen, by age and type of invitation	*	*	*	*	*	*
Number of women screened by age and type of invitation	*	*	*	*	*	*
Women aged 45 and over screened by type of invitation and outcome	*	*	*	*	*	*
Women aged 50-64 screened by type of invitation and outcome	*	*	*	*	*	*
Women screened by age, region and outcome	*	*	*	*	*	-
Women screened by age and outcome	*	*	*	*	*	-
Women screened by age, region and time since last screen		-	*	*	*	-
Women aged 45 and over screened by region and outcome	-	-	-	-	-	*
Women aged 50-64 screened by region and outcome	-	-	-	-	-	*
Women aged 50-64 screened by Strategic Health Authority and outcome	*	*	-	-	-	-
Women aged 45 and over screened and cancers detected by type of invitation	*	*	*	*	*	*
Women aged 50-64 screened and cancers detected by type of invitation	*	*	*	*	*	-
Women screened and cancers detected by age and region	*	*	*	*	*	-
Women screened and cancers detected by age	*	*	*	*	*	*
Women aged 45 and over screened and cancers detected by region	-	-	-	-	-	*

* = available

- = not available

A1 - Breast Screening Programme continued

Data content	Years					
	2002-03	2001-02	2000-01	1999-00	1998-99	1997-98
Women aged 50-64 screened and cancers detected by region	-	-	-	-	-	*
Women aged 50-64 screened and cancers detected by Strategic Health Authority	*	*	-	-	-	-
Women aged 45 and over, with cancer diagnosed, by size of cancer and age	*	*	*	*	*	*
Women aged 50-64, with cancer diagnosed, by invitation type and size of cancer	*	*	*	*	-	*
Coverage of the target age group 50-64 by Health Authority at 31 March	-	*	*	*	*	*
Coverage of women aged 53-64 by Primary Care Organisation, at 31 March	*	-	-	-	-	-
Estimated uptake of invitation to screen, by women aged 50-64 by unit	*	*	*	*	*	*
Estimated uptake of invitation to screen, by women aged 50-64 by type of invitation and unit	*	*	*	*	*	-
Selected diagnostic and outcome statistics for women aged 50-64 by unit	-	*	-	-	-	-

* = available

- = not available

A2 - Cervical Screening Programme

Data content	Years					
	2002-03	2001-02	2000-01	1999-00	1998-99	1997-98
Summary Table: Statistics on cervical cancer and the NHS cervical screening programme, 1993 to 2003	*	-	-	-	-	-
Test status of women and coverage by age, at 31 March of latest year	*	*	*	*	*	*
Test status and coverage of target age group (25-64) by region, at 31 March of latest year	*	*	*	*	*	*
5 year coverage at 31 March by age, time series	*	*	*	*	*	-
3 year coverage at 31 March by age, time series	*	*	*	*	*	-
Test status of women (1) by age, 31 March of latest year	*	*	-	-	-	-
Test status of women by age, 31 March of latest year	*	*	-	-	-	-
Number of women invited in the year by type of invitation and age	*	*	*	*	*	*
Number of women aged 25-64 invited in the year by type of invitation and region			*	*	*	*
Number of women aged 20-64 invited in the year by type of invitation and region	*	*	-	-	-	-
Number of women tested in the year by type of invitation and age	*	*	*	*	*	*
Number of women aged 25-64 tested in the year by type of invitation and region			*	*	*	*
Number of women aged 20-64 tested in the year by type of invitation and region	*	*	-	-	-	-
Number of women aged 20-64 tested in the year by type of invitation and result	*	-	-	-	-	-
Number of tests in the year by type of invitation and result	*	-	-	-	-	-
Test result by age	*	*	*	*	*	*
Test result of women aged 25-64 by region			*	*	*	*
Test result of women aged 20-64 by region	*	*	-	-	-	-
Coverage of the Target Age Group (25-64) and results of tests by Health Authority	-	*	*	*	*	*
Coverage of the Target Age Group (25-64) and results of tests by Primary Care Organisation	*	-	-	-	-	-
Coverage of the Target Age Group (25-64) by Health Authority	-	*	*	-	-	-
Coverage of the Target Age Group (25-64) by Primary Care Organisation, 2001-02	*	-	-	-	-	-

* = available

- = not available

A2 - Cervical Screening Programme continued

Data content	Years					
	2002-03	2001-02	2000-01	1999-00	1998-99	1997-98
Time from screening to notification of result by region, 2002-03	*	*	*	*	-	-
Recall status by most severe screening result and region	*	*	*	*	-	-
Smears examined by pathology laboratories: Time from receipt of smear to authorisation of report by region	*	*	*	-	-	-
Smears examined by pathology laboratories, by source of smear, result of test and Regional Office Area			*	*	*	*
Smears examined by pathology laboratories, by source of smear, result of test and region	*	*	-	-	-	-
Smears examined by pathology laboratories, by source of smear and result of test	*	*	*	*	*	*
GP & NHS Community Clinic smears examined by pathology laboratories, by result and age of women	*	*	*	*	*	*
GP and NHS Community Clinics smears from women aged 20-64 examined by pathology laboratories, by laboratory and result	*	*	*	*	*	*
Outcome of referrals for smears registered at the laboratory between April - June latest year, by region	-	*	*	*	*	*
Outcome of referrals for smears registered at the laboratory between April - June 2002, by Government Office Region	*	-	-	-	-	-
Women referred to colposcopy, October 2000 - March 2001			*	-	-	-
Women referred to colposcopy, England 2002-03	*	-	-	-	-	-
Women referred to colposcopy: by referral indication, result of screening smear and region	-	*	-	-	-	-
Women referred to colposcopy: Time from referral to first offered appointment, by referral indication and region, 2002-03	*	-	-	-	-	-
Women referred to colposcopy: First Attendances by result of referral, type of procedure			*	-	-	-
Women referred to colposcopy: First Attendance by result of referral, type of procedure and region	*	*	-	-	-	-
Biopsies taken at colposcopy: Time from biopsy until patient informed of result			*	-	-	-
Biopsies taken at colposcopy: Time from biopsy until patient informed of result, by region, (4 months sample)	*	*	-	-	-	-
Biopsies taken at colposcopy: Biopsies by type & Outcome			*	-	-	-
Non-diagnostic biopsies taken at colposcopy, by type, outcome and region, (4 months sample)	*	*	-	-	-	-
Summary colposcopy statistics by colposcopy clinic, England 2002 - 03	*	-	-	-	-	-

* = available

- = not available

A3 - NHS Immunisations Statistics

Data content	Years					
	2002-03	2001-02	2000-01	1999-00	1998-99	1997-98
Completed primary courses: percentage of children immunised by their first birthday, time series	*	*	*	*	*	*
Completed primary courses: percentage of children immunised by their second birthday, time series	*	*	*	*	*	*
Completed primary courses: percentage of children immunised by their fifth birthday, time series	*	-	-	-	-	-
Completed primary courses: percentage of children immunised by their second birthday, by region, time series	-	*	*	*	*	*
Completed primary courses: percentage of children immunised by their second birthday, by government office region, 1997-98 to 2002-03	*	-	-	-	-	-
Completed primary courses by age	-	-	-	-	*	*
Completed primary courses, time series starting from 1988-89 ending latest years	-	-	-	-	*	*
Reinforcing doses by age	-	-	-	-	*	*
Reinforcing doses by age, time series starting from 1988-89 ending latest years	-	-	-	-	*	*
Tuberculin skin tests by age and result and BCG vaccinations by age	*	*	*	*	*	*
Tuberculin skin tests by result and BCG vaccinations, time series	*	*	*	*	*	*
Percentage of children immunised by their 2nd birthday, by Health Authority	-	*	*	*	*	*
Percentage of children immunised by their 2nd birthday, by provider	*	-	-	-	-	-
Percentage of children immunised by their 5th birthday, by Health Authority	-	*	*	*	-	-
Percentage of children immunised by their 5th birthday, by provider	*	-	-	-	-	-
Number of children immunised, selected categories, by provider	-	-	*	*	*	*
Number of children receiving BCG vaccinations and reinforcing doses given to school leavers by provider	*	*	-	-	-	-

***** = available

\- = not available

Appendix 4: EQ-5D Instrument

- **Tick one box for each group of statements.**

Mobility

I have no problems in walking about	❑
I have some problems in walking about	❑
I am confined to bed	❑

Self-Care

I have no problems with self-care	❑
I have some problems washing or dressing myself	❑
I am unable to wash or dress myself	❑

Usual Activities

I have no problems with performing my usual activities (e.g. work, study, housework, family or leisure activities)	❑
I have some problems with performing my usual activities	❑
I am unable to perform my usual activities	❑

Pain/Discomfort

I have no pain or discomfort	❑
I have moderate pain or discomfort	❑
I have extreme pain or discomfort	❑

Anxiety/Depression

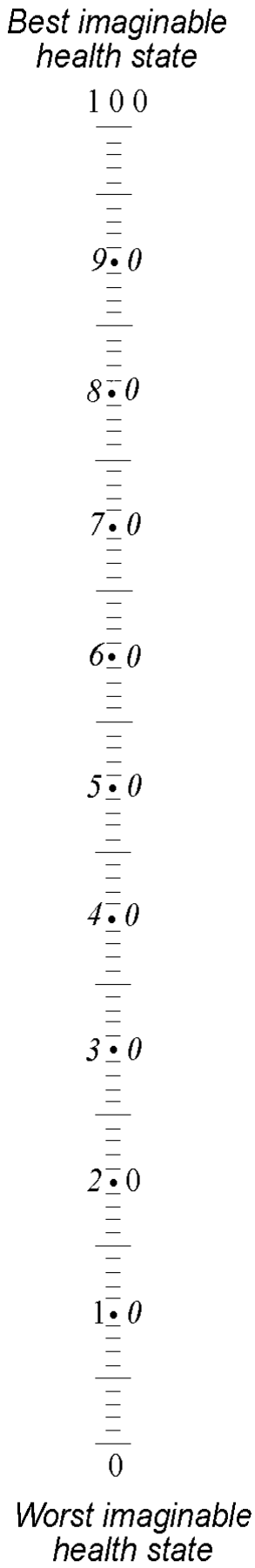
I am not anxious or depressed	❑
I am moderately anxious or depressed	❑
I am extremely anxious or depressed	❑

Compared with my general level of health over the past 12 months, my health state today is:

Better	❑
Much the same	❑
Worse	❑

- **Please indicate on this scale how good or bad your own health is today.**
- **The best health state you can imagine is marked 100 and the worst health state you can imagine is marked 0.**
- **Please draw a line from the box below to the point on the scale that indicates how good or bad your health is today.**

Your own health state today



Appendix 5: EQ-5D Studies

EQ-5D: Study 1

Study number	Source	Authors	Title	Journal	Date	Study description
1	EuroQol Group proceedings	Connor-Spady B, Mintz A, Mallon C et al	Responsiveness of generic health status measures as assessed In patients with rheumatoid arthritis receiving Infliximab	Proceedings of 19th Plenary meeting of EuroQol Group, York	2002	Comparison of a number of health status measures in a group of patients with RA to assess their responsiveness when used in clinical practice
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
		1999-2001	Patients attending rheumatology clinics at University of Alberta (Canada) with persistent active disease (group 2 in the paper). All receiving Methotrexate for at least 6 months.	N=84 at baseline, N=77 at 14 week follow up, but tables give N=60	Not reported	Not reported
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Rheumatoid arthritis	Infliximab	mean EQ-5D index (UK weights)	on day of appointment before first infusion with Infliximab	0.43 (sd 0.30) N=60	14 weeks follow p at time of 4th infusion	0.63 (sd not reported)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.026		

EQ-5D: Study 2

Study number	Source	Authors	Title	Journal	Date	Study description
2	EuroQol Group proceedings	Pickard S, Johnson J, Feeny D	Responsiveness of generic health-related quality of life measures in stroke	Proceedings of 19th Plenary meeting of EuroQol Group, York	2002	Examination of the construct validity of the EQ-5D, SF-36 and HUI in stroke patients
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
		1999-2000	Patients 2-3 weeks post ischaemic stroke confirmed via CT, MRI, EKG scan, at the University of Alberta or Royal Alexandra Hospital, Edmonton (Canada). Patients not eligible if suffered cognitive impairment, aphasia or if life expectancy judged to be < 6 mths	124 patients enrolled, 18 dropped out, 98 completed six month follow up. Data is for N=98	not reported	mean age 67yrs (sd 15)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Ischaemic stroke	natural history study of HRQOL in stroke ~ treatment as usual	mean EQ-5Dindex and EQ-5Dvas (UK weights)	2-3 weeks post stroke, prior to discharge	EQ-5Dindex 0.31 (0.38) N=98	six month follow up, post discharge	EQ-5Dindex 0.62 (0.33) N=98
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.025		

EQ-5D: Study 3

Study number	Source	Authors	Title	Journal	Date	Study description
3	EuroQol Group proceedings	Kind P, Palmer S, Hurst N, Boyd T, Corson M	Generic and condition specific measurement: comparison of EQ-5D and three measures of mental health status	EuroQol Group Plenary Proceedings, Hannover	1998	data taken from 4 different studies to compare the performance of EQ-5D with condition specific health status measures. Data used here taken from Studies A and B in the paper. Study A= trial of 8week psychodynamic interpersonal therapy v usual treatment for patients with psychiatric disorder. Study B= natural history study of HRQOL in rheumatoid arthritis
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
		Not reported	Study A=patients with non-psychotic psychiatric symptoms who had not been responsive to conventional treatment from secondary psychiatric services for at least 6 mths (Manchester); Study B= patients with RA attending outpatient clinics (Edinburgh, Lothian and Fife)	study A=N=106 baseline, 91 at 6 mth follow up; Study B=N=231 baseline, 223 at 3 mth follow up	Study A=63% female; Study B=19% female(??)	Study A mean age 41 yrs (sd 10); Study B mean age 56 yrs (sd 14)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Study A non psychotic psychiatric disorder; Study B=rheumatoid arthritis	Study A=brief psychodynamic interpersonal therapy/treatment as usual; Study B=treatment as usual	Mean EQ-5Dindex and EQ-5Dvas (UK weights)	Study A=after at least 6 months of treatment as usual; Study B=at routine outpatient appointment	Study A EQ-5Dindex 0.36 (0.32); Study B EQ-5Dindex 0.38 (0.37)	Study A= 6 mths; Study B=3 mths	Study A EQ -5Dindex 0.41 (0.32); Study B EQ-5Dindex 0.43 (0.36)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.024		

EQ-5D: Study 4

Study number	Source	Authors	Title	Journal	Date	Study description
4	EuroQol Group proceedings	Longworth L, Ratcliffe J, Young T, Bryan S	A comparison of EQ-5D and SF-6D single index in assessing the health-related quality of life of liver transplant patients	Proceedings of 18th Plenary meeting of EuroQol Group, Copenhagen	2001	Comparison of EQ-5D and SF 6D scores in patients listed for liver transplantation
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
		1995-1996	All adult patients listed for liver transplantation at all 6 NHS designated transplant centres	N=585 eligible for study. Two data points for n=152	49% female	49 yrs (sd 12)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Liver disease requiring transplant	Liver transplantation	mean EQ-5Dindex score (UK weights)	at time of listing for transplantation	EQ-5Dindex 0.53 (sd ?) N=152	24 months post transplant	EQ-5Dindex 0.59 (sd ?) N=152
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.023		

EQ-5D: Study 5

Study number	Source	Authors	Title	Journal	Date	Study description
5	EuroQol Group proceedings	Peerenboom L, Krabbe PFM	Responsiveness of Eq-5D: HRQOL outcomes in a clinical study on surgical treatments of colorectal hepatic metastases	Proceedings of 18th Plenary meeting of EuroQol Group, Copenhagen	2001	Examination of responsiveness to change of Eq-5D in a clinical study of treatment of colorectal liver metastases
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
		1999-2001	Patients requiring surgical resection of colorectal liver metastases with/without local ablative surgery, Nijmegen NL. 3 groups - resectable with/out local ablation; local ablation only; irresectable	resectable w/out ablation N=28; local ablation only N=10; irresectable N=13	not reported	not reported
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
metastatic disease confined to liver	resection w/out local ablative therapy	mean EQ-5Dindex and EQ-5Dvas (UK weights)	pre-operatively (exact time not reported)	resection EQ-5Dindex 0.8 (sd ?); local ablation 0.9 (sd ?); irresectable 0.7 (sd ?)	18 days post surgery	resectable EQ-5Dindex 0.7 (sd ?); local ablation 0.7 (sd ?); irresectable 0.5 (sd ?)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.022		

EQ-5D: Study 6

Study number	Source	Authors	Title	Journal	Date	Study description
6	Survey	Wilson A, Parker H, Wynn A, Jagger C, Spiers N, Jones J, Parker G	Randomised controlled trial of effectiveness of Leicester hospital at home scheme compared with hospital care	BMJ 319: 1542-6	1999	Pragmatic RCT. GPs referrals randomised to Leicester hospital at scheme or the city's three acute hospitals
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	Leicester	November 1991 to May 1997	199 consecutive patients referred to hospital at home by the GP and assessed as being suitable for admission - 102 to hospital at home and 97 to hospital.	199	78 m/121 w	Age range 33 -102
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Various. Largest categories were cardiovascular and respiratory	Referral to hospital at home or inpatient care	Median EuroQol reported	3 days after randomisation	not reported	Two weeks after randomisation	Hospital at home 0.59 (0.15 - 0.78) Hospital 0.56 (0.19-0.73)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
3 months after randomisation	Hospital at home 0.64 Hospital 0.63			Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.021		

EQ-5D: Study 7

Study number	Source	Authors	Title	Journal	Date	Study description
7	Survey	Ascione R, Reeves BC, Taylor FC, Seehra HK, Angelini G	Beating heart against cardioplegic arrest studies (BHACAS 1 and 2): quality of life at mid-term follow-up in two randomised controlled trials	European Heart Journal 25: 765- 770	2004	Compare generic and disease specific QoL two to four years after surgery in participants in two RCTs of OPCAB vs CABG-CPB
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Cross-sectional	Bristol	2001	Patients who had been recruited for RCTs following heart disease in 1997-1999.Of the original 401 patients, 22 had died and and 51 did not respond to invitation	328	270 m/58 w	Mean age CABG-CPB 61.4, OPCAB 63.1
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Heart disease	Patients who had been allocated either to CABG-CPB or OPCAB surgery	Mean Index and VAS reported	2, 3 or 4 years after surgery	Index: CABG-CPB n = 151, 0.82 (sd 0.25); OPCAB n = 161 0.81 (sd 0.24). VAS CABG-CPB n = 151 77.0 (sd 19.1); OPCAB n = 165 76.0 (sd 16.1)		
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.020		

EQ-5D: Study 8

Study number	Source	Authors	Title	Journal	Date	Study description
8	Survey	Goodacre S, Nicholl J, Dixon S, Cross E, Angelini K, Arnold J et al	Randomised controlled trial and economic evaluation of a chest pain observation unit compared with routine care	BMJ, online doi:10.1136/bmj.37956.664236.EE (published 14 January 2004)	2004	Cluster randomized controlled trial with 442 days randomised to the chest pain observation unit or routine care
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	Sheffield	Feb 2001 - May 2002	972 patients with acute, undifferentiated chest pain (479 attending on days when care was delivered in the chest pain obs unit, 493 on days of routine care).	972	622 m/350 w	Mean age CPOU 49.4, Routine care 49.6
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Chest pain	Chest pain observation unit or routine care	Mean VAS reported ? They refer to it as health utility in their table	2 days after admission to study	Chest pain unit Mean VAS 79.8, Routine care 75.7	One month	Chest pain unit 77.9, Routine care 71.86
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
6 months	Chest pain unit 79.0; routine care 76.3			Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.019		

EQ-5D: Study 9

Study number	Source	Authors	Title	Journal	Date	Study description
9	Survey	Goodacre, Mason S, Arnold J, Angelini K	Psychological Morbidity and Health-related Quality of Life of Patients assessed on a chest pain observation unit	Annals of Emergency Medicine; 38: 369-376	2001	To measure psychological morbidity and health related quality of life among patients attending hospital with acute chest pain, at presentation and one month later.
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	Sheffield	October 1999 - April 2000	All patients who were admitted to CPOU (unless communication problems). 168 patients completed baseline and 119 follow up questionnaires	168	60% men	Mean age, baseline only 49, baseline and follow-up 54.
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Chest pain	Chest pain observation unit	Mean Index reported	On presentation	Mean Index 0.63 n = 166	One month	Mean index 0.69, n = 110
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.018		

EQ-5D: Study 10

Study number	Source	Authors	Title	Journal	Date	Study description
10	Survey	Giannoudis PV, Nicolopoulos C, Dinopoulos H, NgA, Adedapo S, Kind P	The impact of lower leg compartment syndrome on health related quality of life	Injury, International Journal of the care of the Injured; 33: 117-121	2002	To assess long-term impact of compartment syndrome on quality of life. Questionnaire survey of patients who had been treated for this problem between 1993 and 1998
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Cross-sectional	Britain (Leeds ?)	?	30 patients who underwent fasciotomies. One excluded from final analysis due to complex fracture. Also a control group of patients on data base but without compartment syndrome	29. Control 33	25 m/4 w control group 23 m/7w	Mean age was 35 (range 19 - 65). Control group 39 (range 18 - 68).
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Lower leg injury	Fasciotomy within 12 hours of injury	Mean VAS and Index for patients with and without impact on function, skin graft and problem with appearance	One year or more after surgery	Impact on function Yes mean VAS 71.99 No Mean VAS 95.1, Yes Mean Index 0.619 No Mean Index 0.946. Skin graft Yes mean vas 74.1 No 89.0 Yea mean Index 0.619 No 0.946. Appearance some problem Mean VAS 74.2 Not a problem 93.5, Appearance some problems Mean Index 0.631, Not a problem 0.931		
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.017		

EQ-5D: Study 11

Study number	Source	Authors	Title	Journal	Date	Study description
11	Survey	Gopal S, Giannoudis PV, Murray A, Matthews SJ, Smith RM	The functional outcome of sever, open tibial fractures managed with early fixation and flap coverage	Journal of Bone Joint Surgery: 85 - B	In press, 2004	Outcome and functional status of patient with severe open tibial fractures after and 'fix and flap' regime
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Cross-sectional	Leeds	?	33 patients who had been treated between 1996 and 2000	33	27m/6w	Mean age of the adult group was 48 years, range 19-79
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Lower leg injury	Assessment following surgery. Mean follow-up time of final review was 46 months (15-80)	Mean VAS		Overall mean for adults in survey was 68		
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.016		

EQ-5D: Study 12

Study number	Source	Authors	Title	Journal	Date	Study description
12	Survey	Barlow JH, Turner AP and Wright CC	A randomized controlled study of the Arthritis Self-Management Programme in the UK	Health Education Research;15, 6: 665-680	2000	To determine whether the Arthritis Self-Management programme (ASMP) improves perceptions of control, health behaviour and health status, and changes use of health care resources
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK	?	GP attenders with diagnosis of arthritis invited to take part in ASMP, then randomized into an intervention group or control group (on 4 month waiting list)	544: Intervention Group = 311 Control Group = 233	w = 85% Intervention Group and 83% Control Group	Mean age Intervention group = 57.3 and Control Group 59.1
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Arthritis	Intervention Group received Arthritis Self-Management Programme ie six weekly sessions each lasting approx 2h, delivered by pairs of lay leaders trained by Arthritis Care	Mean Index and VAS reported. EQ-5D only used with a subsample. Scores available for Intervention group n = 86 and Control Group n = 78	After randomization	Mean Index Intervention Group = 0.43 Control Group = 0.44 Mean VAS Intervention Group = 56.78 Control Group = 57.87	4 month follow-up	Mean Index Intervention group = 0.47 Control group = 0.45 Mean VAS Intervention Group = 57.12 Control Group = 59.04 Differences in changes between baseline and 1st follow-up = n/s
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
12 month follow-up Intervention Group only	No scores given but "There was a significant improvement on the EuroQol VAS (p = 0.0007)			Changes in scores for PIN group were stat significant at 6 weeks p = 0.009 and at 6 months p = 0.015		

EQ-5D: Study 13

Study number	Source	Authors	Title	Journal	Date	Study description
13	Survey	Williams NH, Wilkinson C, Russell I, Edwards RT, Hibbs R, Linck P, Muntz R	Randomized osteopathic manipulations study (ROMANS): pragmatic trial for spinal pain in primary care	Family Practice 2003; 20: 662-669	2003	To assess the effectiveness and health care costs of a practice-based osteopathy clinic for subacute spinal pain. RCT carried out in primary care osteopathy clinic accepting referrals from 14 neighbouring practices in North West Wales
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - Wales	September 1997 - March 2001	Patients aged between 16 and 65 presenting to 14 general practices with mechanical pain in the neck or upper or lower back of 2-12 weeks duration, either the first episode or a recurrence	201 randomised to usual GP care = 109 or ostepathic treatment = 92	n/I	n/I
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Spinal pain	Usual care or usual care + referral to osteopathy clinic for three or four sessions of spinal manipulation and advice	Mean Index and Vas	After randomization	Mean Index Intervention Group = 0.56 Control Group = 0.50 Mean VAS Intervention Group = 58.9 Control Group = 61.3	2 month follow-up[	Mean Index Intervention group = 0.67 Control group = 0.56 Mean VAS Intervention Group = 69.5 Control Group = 66.1
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
6 month follow-up	Mean Index Intervention group = 0.66 Control group = 0.60 Mean VAS Intervention Group = 69.1 Control Group = 66.4			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.014		

EQ-5D: Study 14

Study number	Source	Authors	Title	Journal	Date	Study description
14	Survey	Kaur S, Lambourne P and Ashaye, K	Day hospital care for older people with mental health problems	Health & Ageing, The Clinician 2003:4 : ii-iii	2003	To assess the impact of Care Programme Approach meetings and day hospital care on the **informal carers** who lived with patients
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Cross-sectional	UK - North Hertfordshire	n/i	Carers of older patients with mental health problems attending a day hospital	30 carers	15 m/15 w	Mean age of 71.1
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Carers of older patients with mental health problems	n/a	Mean Index and VAS		Mean Index = 0.70 and mean VAS 67.36		
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.013		

EQ-5D: Study 15

Study number	Source	Authors	Title	Journal	Date	Study description
15	Survey	Lacey EA, Walters SJ	Continuing inequality: gender and social class influences on self perceived health after a heart attack	Journal of Epidemiol Community health 2003; 57: 622-627	2003	Longitudinal survey design to investigate the effect of social class and gender on self-perceived health status for thos recovering from an acute myocardial infarction. Community bases study in a city in the north of England
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - city in north of England	1998-9	People discharged from hospital after acute myocardial infarction	229 people	166 m/ 55 w	Mean for whole sample 62.4. Mean m = 61.8, mean women = 64.4
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
People discharged from hospital after acute myocardial infarction	none	Mean Index, reported for whole sample and various subgroups (gender, educational qualifications, car access)	Six weeks after discharge	Mean index = 0.68	Six months after discharge	not reported
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
One year after discharge	Mean index = 0.72			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.012		

EQ-5D: Study 16

Study number	Source	Authors	Title	Journal	Date	Study description
16	Survey	Holland R, Smith, RD, Harvey I, Swift L and Lenaghan E.	Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL	Health Economics (in press) Published online in Wiley InterScience. COI:10.1002/hec.858	2004	Comparison of EQ-5D and AQoL as part of a randomised controlled trial to investigate the cost-effectiveness of home based medication review by pharmacists for patients after over 79 years
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - Norfolk	n/i	Patients aged over 79 years,discharged from hospital, taking two or more medications each day, admitted as an emergency and returning to their own home or warden controlled accommodation	145 at baseline, four excluded from trial after baseline and 17 withdrew at follow-up	62 m / 83 m	Mean = 84.7
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Aged over 79 years	Home-based medication rewiew by pharmacists	Mean Index reported for whole sample and various subgroups	At recruitment 'shortly before discharge' Questionnaires completed with help of a recruiter	Mean Index 0.61	Three months post recruitment, by postal survey	not reported
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
Six months post recruitment, by postal survey	Mean Index = 0.45			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.011		

EQ-5D: Study 17

Study number	Source	Authors	Title	Journal	Date	Study description
17	Survey	Cohn, TJ, Foster JH, Peters TJ	Sequential studies of sleep disturbance and quality of life in abstaining alcoholics	Addiction Biology 2003; 8: 455-462	2003	Longitudinal survey to monitor sleep quality and quality of life in abstaining alcoholics Qustionnaires completed at baseline and for 12 weeks at monthly intervals
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - London	n/i	Inpatients at a voluntary sector residential alcohol treatment centre	57	38 m / 19 w	Age range = 23 - 69 Mean of 42 for both men and women
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Abstaining alcoholics	Inpatients at a voluntary sector residential alcohol treatment centre, abstaining from alcohol	Mean VAS reported	At recruitment after a minimum 7-day residnece in the unit	54.1	4 weeks after recruitment	67.5
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
8 weeks after recruitment	77.4	12 weeks after recruitment	81.6	Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.010		

EQ-5D: Study 18

Study number	Source	Authors	Title	Journal	Date	Study description
18	Survey	Millward LM, Kelly P, Deacon A, Senior V and Peters TJ	Self-rated psychosocial consequences and quality of life in the acute prophyrias	J Inherit. Metab. Dis 2001; 24: 733-747	2001	Postal survey of patients with porphyria to investigate prevalence of psychosocial symptoms andperceived effects or porphyria on quality of life and patient experience
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Cross-sectional	UK - patients referred to a London service	n/i	116 patients aged 18 years and over who had been referred for clinical management or laboratory diagnosis to the London Supraregional Assey Sercie Centre for Porphyria	Postal survey sent to 116 patients, usable replies received from 81 patients	22 m / 53 w	Mean age was 43
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Porphyria	n/a	Mean Index and VAS by age and gender, compared with normative data	Questionnaires sent to any patients over 18 year who had attended the clinic in the previous 10 years	Mean Index ranged from 0.10 for 70-79 year old males to 0.86 for 20-29 year old females. Mean VAS ranged from 35.0 for 790-79 year old males to 77.5 for 30-39 year old males. Patients with manifest symptomology had VAS scores of 0.71 compared with those with latent symptomology of 0.76.		
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.009		

EQ-5D: Study 19

Study number	Source	Authors	Title	Journal	Date	Study description
19	Lit Search	Steadman, J, Donaldson N, Klara L	A Randomized controlled Trial of an Enhanced Balance Training Program to Improve Mobility and Reduce Falls in Elderly Patients	Journal of the American Geriatric Society 2003; 51:847-852	2003	Evaluation of effectiveness of blance training program. Prospective, single-blind RCT
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK- South London District General Hospital	n/i	198 subjects aged 60 and older allocated to intervention group for enghanced blance training (EBT) or to the control group receiving conventional physiotherapy (CT)	198 patients Intervention Group = 96 and Control Group = 102	80% w	Mean age was 82
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Elderly people with blance problems (Berger Balance Scale score of less than 45)	Enhanced Balance Training (EBT) or conventional physiotherapy (CT)	Mean VAS	On randomisation	EBT group - mean VAS 57.8 +/- 19.7 CT group - mean VAS 59.4 +/ 17.2	6 weeks later	EBT group - mean VAS 65.1 +/- 19.6 CT group - mean VAS 64.9 +/ 17.3
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
12 weeks	EBT group - mean VAS 65.1 +/- 17.7 CT group - mean VAS 65.7 +/ 16.9	24 weeks	EBT group - mean VAS 64.4 +/- 19.9 CT group - mean VAS 64.5 +/ 17.4	Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.008		

EQ-5D: Study 20

Study number	Source	Authors	Title	Journal	Date	Study description
20	Lit Search	Manca A, Sculpher MJ, Ward K, Hilton P	A cost-utility analysis of tension-free vaginal tape *versus* colposuspension for primary urodynamic stress incontinence	BJOG: an International Journal of Obstetrics and Gynaecology 2003; 110: 255-262	2003	Cost-utility analysis alongside a multcentre randomised comparative trial
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK and Ireland, gynaecology and urology departments in 14 centres	n/i	Women with urodynamic stress incontinence	344 patients recruited - colposuspenion = 169 and tension-free vaginal tape = 175. 34 dropped out	all female	not reported in this article
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Women with urodynamic stress incontinence	Either colpsuspension or tension-free vaginal tape	Mean Index	randomisation	T-f vag tape 0.78 [0.71 - 0.92] Colposusp. 0.79 [0.71 - 0.92]	6 weeks	T-f vag tape 0.79 [0.71 - 0.92] Colposusp. 0.75 [0.69 - 0.88]
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
6 months	T-f vag tape 0.81 [0.73 - 0.92] Colposusp. 0.79 [0.73 - 0.92]			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.007		

EQ-5D: Study 21

Study number	Source	Authors	Title	Journal	Date	Study description
21	Lit search	Round A, Crabb T, Buckingham K, Mejzner R, Pearce V, Ayres R, Weeks C and Hamilton W.	Six month outcomes after emergency admission of elderly patients to a community or a district general hospital	Family Practice 2004; 21, No. 2. Doi: 10.1093/fampra/cmh212	2004	Prospective cohort study to compare patient based outcomes following emergency admission to a community or district general hospital
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - Devon, one DGH and five CHS	November 1999 to November 2000	376 patients aged > 70 years with an acute illness requiring hospital admission, but whose condition could have been treated in either hospital setting	376 at baseline, but only 254 followed up at 6 month stage (136 CH, 118 DGH)	of 254 who were followed up at 6 month stage: 126 m/ 128 = f	Median age = 82 (76-88)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Elderly patients with an acute illness requring hospital admission	One of two settings, community or district general hospital	Mean VAS, Median Index Mean Change in VAS and Index over six months	Within 48 hours of admission	Mean VAS DGH group 53.9 (50.7-57.1) CH group 50.5 (47.5-53.6) Median INDEX DGH group 0.36 (0.07-0.69) CH group 0.26 (0.0005-0.69)	6 months later	
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.006		

EQ-5D: Study 22

Study number	Source	Authors	Title	Journal	Date	Study description
22	Lit Search	Jarman B, Hurwitz B, Cook A, Bajekal M, Lee A	Effects of community based nurses specialising in Parkinson's disease on health outcome and costs: randomised controlled trial	BMJ 2002: 324; 1072 - doi: 10.1136/bmj.324.7345.1072	2002	Two year RCT to determine the effects of community based nurses specialising in Parkinsons's disease on health outcomes and healthcare costs. Patients randomised to nurse specialist group (56% of patients) or the control group
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - 9 randomly selected health authority areas of England	1995	Patients aged 18 and over with Parkinsons' disease identified by 438 general practices	1836 patients recruited. After two years 315 (17.3%) had died	1044 m/ 792 f	Less than 70 = 610, 70-77 = 649, More than 77 = 577
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Parkinsons' disease	Either nurse specialist care or standard care (control group)	Mean EQ-5D INDEX	randomisation	Mean INDEX 0.47 (0.35)	Two years later	Mean INDEX Nurse Group 0.37 |(0.35) Control Group 0.39 (0.35)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.005		

EQ-5D: Study 23

Study number	Source	Authors	Title	Journal	Date	Study description
23	Lit Search	Allsup SJ, Gosney MA	Anxiety and depression in an older research population and their impact on clinical outcomes in a randomised controlled trial	Postgrad Medical Journal 2002; 78: 674-677	2002	RCT investigating the cost benefits of influencza vaccination over six months. Participants were randomised in a ration 3:1 to receive either influenza vaccination or placebo. As part of this study the impact of anxiety and depression on results was investigated and reported here.
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK	Recruitment commenced October 1999	Fit and healthy, independent living 65-74 year olds not previously requiring influenza vaccination	729 individuals	not reported	not reported
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
n/a	Influenza vaccination or placebo	Mean VAS scores compared according to whether person had high or low scores on HADS	After randomisation but before vaccination	Mean (SD) VAS for low HADS group 84..5 (14.4) for high HADS group 63.1 (17.4) Difference = sign. (p<0.001)	2 months after injection	Mean (SD) VAS for low HADS group 83.2 (12.7) for high HADS group 58.9 (18.4) Difference = sign. (p<0.001)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
4 months after injection	Mean (SD) VAS for low HADS group 83.7 (11.9) for high HADS group 57.6 (19.9) Difference = sign. (p<0.001)	6 months after injection	Mean (SD) VAS for low HADS group 84.0 (12.7) for high HADS group 53.3 (19.9) Difference = sign. (p<0.001)	Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.004		

EQ-5D: Study 24

Study number	Source	Authors	Title	Journal	Date	Study description
24	Lit Search	Kendrick D, Field K, Bentley E, Kerslake R, Miller P and Pringle M	Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial	BMJ 2001; 322: 400-405	2001	Randomised unblinded controlled trial to investigate the impact of radiographyof lumbar spine on improved clinical outcomes or satisfaction with care
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - 52 general practices in England	Recruitment between November 1995 and January 1999	Patients with low back pain of a median duration of 10 weeks	421 patients	174 m / 247 f	Median age 39 (31 - 46)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Low back pain	Intervention group received usual care and radiograph of the lumbar spain at their local hospital. Control group received the usual care from their doctor, including radiography if the doctor considered it to be clinically necessary	Median EQ-5D INDEX and VAS (?called health status score)	Before randomisation	Median EQ-5D INDEX (interquartile range) Intervention Group = 0.69 (0.62 - 0.76) Control group = 0.69 (0.62-0.76) Median VAS (interquartile range) Intervention group 70 (50-80) Control group 70 (50-80)	3 months after randomis-ation	Median EQ-5D INDEX (interquartile range) Intervention Group = 0.80 (0.69 - 0.88) Control group = 0.80 (0.69-0.91) Median VAS (interquartile range) Intervention group 75 (60-90) Control group 80 (70-90)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
9 months after randomis-ation	Median EQ-5D INDEX (interquartile range) Intervention Group = 0.80 (0.69 - 1.00) Control group = 0.80 (0.73-1.00) Median VAS (interquartile range) Intervention group 80 (60-90) Control group 80 (70-90)			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.003		

EQ-5D: Study 25

Study number	Source	Authors	Title	Journal	Date	Study description
25	Lit Search	Sculpher M, Manca A, Abbott J, Fountain J, Mason S and Garry R	Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial	BMJ (2004); 328; 134 doi: 10.1136/bmj.37942.601331.EE	2004	To assess cost effectiveness of laparoscopic hysterectomy compared with conventional hysterectomy (abdominal or vaginal)
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK	?	Women requiring a hysterectomy for reasons other than malignancy	1346 women	all female	not reported
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Gynaecology	Laparoscopic hysterectomy or standard hysterectomy (either abdominal or conventional accordin gwyhich was most appropriate) giving two trials vaginal versus laparoscopic and abdominal versus laparoscopic hysterectomy 2;1 randomisation in favour of the laparoscopy	Mean EQ-5D INDEX	?	Mean EQ-5D INDEX: Vaginal Trial, Lap group 0.75, vag group 0.76. Abdominal trial Lap group 0.72 abdom group 0.69	6 weeks	Mean EQ-5D INDEX: Vaginal Trial, Lap group 0.88, vag group 0.85. Abdominal trial Lap group 0.83 abdom group 0.83
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
4 months	Mean EQ-5D INDEX: Vaginal Trial, Lap group 0.91, vag group 0.92. Abdominal trial Lap group 0.89 abdom group 0.87	One year	Mean EQ-5D INDEX: Vaginal Trial, Lap group 0.92, vag group 0.92. Abdominal trial Lap group 0.90 abdom group 0.89	Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.002		

EQ-5D: Study 26

Study number	Source	Authors	Title	Journal	Date	Study description
26	Lit Search	Jenkinson C, Stradling J, Petersen S	Comparison of three measures of quality of life outcomes in the evaluation of continuous positive airways pressure therapy for sleep apnoea	Journal of Sleep Research (1997); 6: 199-204	1997	To compare different measures of quality of life in patients treated with nasal continuous positive airways pressure
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - Oxford Sleep Clinic	Recruitment from April 1995 to February 1996	Men attending for a therapeutic assessment of NCPAP therapy	108 patients	all male	Mean age was 50.0 years (range 28 - 74)
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Sleep disorders	Nasal continuous positive airways pressure	Mean EQ-5D INDEX	Before treatment	Mean (SD) EQ-5D INDEX: 0.79 (0.21)	5 - 7 weeks after the assessment of NCPAP had started	Mean (SD) EQ-5D INDEX:0.84 (0.25)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.001		

EQ-5D: Study 27

Study number	Source	Authors	Title	Journal	Date	Study description
27	Lit Search	Jenkinson C, Stradling J Petersen S [note this is v. similar to one above, but not quite the same]	How should we evaluation health status? A comparison of three methods in patients presenting with obstructive sleep apnoea	Quality of Life Research (1998); 7:95-100	1998	To compare three approaches to the measure of patient reported health status
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - Oxford Sleep Clinic	March 1995 to October 1996	Mean attending for assessment for CPAP	108	all male	Mean age of 49, range 28 - 72
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Sleep disorders	Continuous positivie airways pressure therapy	Mean EQ-5D INDEX and VAS	On assessment	Mean EQ-5D INDEX: 0.78(0.22) Mean EQ-5D VAS: 66.57 (18.91)	3 months later	Mean EQ-5D INDEX: 0.83(0.22) Mean EQ-5D VAS: 71.72 (18.12)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
				Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.000		

EQ-5D: Study 28

Study number	Source	Authors	Title	Journal	Date	Study description
28	Lit Search	Jenkinson C, Gray A, Doll H, Lwrence K, Keoghane S, Layte R	Evaluation of Index and Profile Measures of Health Status in a Radnomized Controlled Trial	Medical Care (1997); 35:1109-1118	1997	To compare two generic measure of health status with disease-specific measures in a randomized controlled trial of transurethral resection of the prostate with laser vaporization prostatectomy for benign prostate hypertrophy
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK - patients recruited at Churchill Hospital in Oxford	?	Men attending hopsital for surgical treatment for benign prostate hypertrophy	?	all male	Average age was 70 years, range 48-93
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Benign prostate hypertrophy	Transurethral resection of the prostate (TURP) or lasor vaporization prostatectomy	Mean EQ-5D INDEX and VAS	Preoperative	EQ-5D INDEX: TURP group 0.81(0.18), Laser group 0.81(0.18) EQ-5D VAS: TURP group 78.3 (13.2) Laster group 75.8(17.1)	3 months later	EQ-5D INDEX: TURP group 0.85(0.17), Laser group 0.85(0.20) EQ-5D VAS: TURP group 79.9 (16.3) Laster group 74.2(19.5)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
1 year	EQ-5D INDEX: TURP group 0.82(0.22), Laser group 0.82(0.21) EQ-5D VAS: TURP group 77.2 (16.9) Laster group 76.5(18.1)			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.001		

EQ-5D: Study 29

Study number	Source	Authors	Title	Journal	Date	Study description
29	Lit Search	Hurley DA, Minder PH, McDonough SM, Walsh DM, Moore AP, Baxter DG	Interferential Therapy Electrode Placement Technique in Acute Low Back Pain: A Preliminary Investigation	Archives of Physical Medicine and Rehabilitation (2001); 82: 485-93	2001	To determine the eficacy of interferential therapy (IFT) electrode placement echnique compared with a control treat of subjects with acute low back pain. Single-blind RCT with a 3-month follow-up
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK	?	Patients with acute lower back pain, referred by GP for physiotherapy treatment at a large acute care, general hospital	59 + control group of 20	46.7% male	Range 19-62 years
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Acute lower back pain	IFT painful area and 'The Black Book, IFT spinal nerve and 'The Black Book OR Control group 'The Black book' only	Median EQ-5D INDEX	Before commencing treatment	Median EQ-5D INDEX Painful area group 0.69, Spinal Nerve group 0.76, Control group 0.69	Discharge	Median EQ-5D INDEX Painful area group 0.80, Spinal Nerve group 0.79, Control group 0.93
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
3 months later	Median EQ-5D INDEX Painful area group 0.80, Spinal Nerve group 0.80, Control group 1.0			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.002		

EQ-5D: Study 30

Study number	Source	Authors	Title	Journal	Date	Study description
30	Lit Search	Durkin MT, Lurton EPL, Wijesinghe LD, Scott CJS, Berridge DC	Long Saphenous Vein Stripping and Quality of Life - a Randomised Trial	European Journal of Vascular and Endovascular Surgery (2001); 21: 545-549	2001	To assess the quality of life of patient undergoing sapheno-femoral junction ligation and long saphenous veing stripping using two different techniques, Prospective randomised trial.
Study design	**Geographical location**	**Date of study**	**Study population**	**Sample Size**	**Gender**	**Age**
Longitudinal	UK	?	Patients from the venous outpatient's clinic	80	28 m / 52 f	Men, median age 56, range 22 - 70. Women, median age 41, range 23 - 70.
Diagnostic Group	**Treatment**	**EQ-5D scores**	**Timing of baseline observation**	**Baseline EQ-5D**	**1st Follow-up time**	**1st Follow-up EQ-5D**
Varicose veins	PIN stripping (43 patients) or Conventional stripping (37)	Median EQ-5D INDEX	Pre-op	Median EQ-5D INDEX PIN group 0.73 (0.66-0.83) Conventional group 0.8 (0.69-1.0)	6 weeks	Median EQ-5D INDEX PIN group 0.8 (0.73 - 1.0) Conventional group 0.83 (0.69-1.0)
2nd follow-up time	**2nd follow-up EQ-5D**	**3rd follow-up time**	**3rd follow-up EQ-5D**	**Comments**		
6 months	Median EQ-5D INDEX PIN group 1.0 (0.73 - 1.0) Conventional group 1.0 (0.69-1.0)			Changes in scores for PIN group were stat signiciant at 6 weeks p = 0.009 and at 6 months p = 0.003		