

Report of the Office of Health Economics Commission on

NHS Outcomes, Performance and Productivity

March 2008



Acknowledgement

The work of the Office of Health Economics Commission on NHS Outcomes, Performance and Productivity has been funded by unrestricted grants from the Association of the British Pharmaceutical Industry, GlaxoSmithKline plc, Roche and Sanofi Aventis. We thank them for enabling the Commission to do its work and for their support of the Commission's aims.

We have been delighted also to receive expressions of support for our objectives and for our work towards achieving them from: the Association of Chartered Certified Accountants, the National Association of Primary Care, the NHS Alliance, the NHS Confederation and the Royal College of Nursing.

The OHE Commission worked independently, and the report expresses the personal views of the Commission's members. They should not be taken to represent the views of members' employers, or of the Commission's funders and other supporters.

About the Office of Health Economics

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- Commission and undertake research on the economics of health and health care;
- Collect and analyse health and health care data from the UK and other countries;
- Disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by revenue from commissioned research, consultancy and sales of publications.

The research and editorial independence of the OHE is ensured by its Policy Board, chaired by Professor Tony Culyer (University of York).

Commission members:

Peter Smith (Centre for Health Economics, University of York) – Commission Chair
Nick Black (London School of Hygiene and Tropical Medicine)
Roger Boyle (National Director for Heart Disease, Department of Health)
Angela Coulter (Picker Institute)
Nancy Devlin (Department of Economics, City University)
Nigel Edwards (NHS Confederation)
Mike Richards (National Director for Cancer, Department of Health)
Adrian Towse (Office of Health Economics)
Andrew Vallance-Owen (BUPA)

Secretariat:

Deven Chauhan (Office of Health Economics)
Jon Sussex (Office of Health Economics)

CONTENTS

FOREWORD BY PROFESSOR PETER SMITH	4
EXECUTIVE SUMMARY	5
1. INTRODUCTION	13
2. METHOD	23
3. RESEARCH FINDINGS – PATIENT OUTCOMES	28
4. RESEARCH FINDINGS – WIDER BENEFITS	47
5. AGGREGATION ISSUES	55
6. IMPLEMENTATION ISSUES	60
7. CONCLUSIONS AND RECOMMENDATIONS	67
REFERENCES	71
APPENDICES	74

FOREWORD BY PROFESSOR PETER SMITH

Ever since Florence Nightingale, commentators have recognised that health system performance should ultimately be assessed in terms of the outcomes it achieves, expressed most immediately in improvements in the length and quality of life of patients. Furthermore, few would dispute the claim that limited health care resources should in principle be allocated so as to secure the maximum levels of such outcomes. However, in spite of occasional successes and progress, attempts to assess health system performance in terms of outcomes and productivity have until recently been frustrated by severe limitations in the scope, quality and timeliness of data.

That situation is now changing dramatically. The IT revolution is making it feasible to capture a previously unimaginable range of information about the inputs, processes and outcomes of health care. Furthermore, an increasingly challenging public is demanding greatly improved accountability in exchange for the massive levels of taxes it invests in the health system.

It is against this background that in autumn 2006 the Office of Health Economics set up a Commission on NHS Outcomes, Performance and Productivity. A distinguished panel of commissioners was assembled, and OHE very rapidly mobilized inputs from an impressive range of researchers, managers and health care professionals to inform the Commission's deliberations.

Commentators have been pressing for more outcome measurement in the NHS for numerous purposes, such as informing patients about providers and treatments, professional improvement, enhancing research potential, promoting accountability, monitoring clinical performance and allocating resources. To some extent, this cacophony of pressures must appear confusing. However, early on in its work, the Commission judged that a crucial nexus for consideration of many aspects of outcome measurement was the commissioner of health care at local, regional and national levels.

This report is therefore aimed predominantly (though not exclusively) at purchasers of health care within the NHS, and seeks to offer a pragmatic but ambitious path for the rapid development of outcome measurement over the next few years. The intention is to prompt a fast take-up of measurement wherever feasible and to highlight more ambitious possibilities for the future. The Commission believes that improvements in outcome measurement, and the associated ability to measure productivity, are vital instruments for improving the performance of the NHS, expressed ultimately in improved quality of care received by patients and better preventative measures.

Furthermore, the improved measurement will undoubtedly enhance the accountability of the health system to the general public. Only when the NHS is prepared to undergo such objective scrutiny can it hope to secure the unassailable legitimacy it needs for the continued support of voters, taxpayers, patients and carers.

As we were drafting this report, the Department of Health published its "Operating Framework for the NHS in England 2008/09", which requires all NHS providers to collect and report patient reported outcome measures for elective hip and knee replacements, groin hernia surgery and varicose vein procedures from April 2009. We hope that this report will help commissioners and providers to implement this promising initiative, and also to see how outcome measurement across a much broader range of activity can be introduced rapidly into the NHS.

I should like to express my thanks to my fellow commissioners, the many researchers and professionals who contributed to the Commission, to the funders and to the OHE secretariat, who worked very impressively to a very tight timetable. After a long period of gestation, it does now look as if the widespread and routine use of outcome measurement is a concept whose time has arrived. I suspect that in 20 years' time it will seem beyond belief that we struggled for so long to steer our health system in the absence of any clear idea of what it was achieving. It is not too fanciful to hope that the era of 'flying blind' in the NHS is rapidly coming to an end.

EXECUTIVE SUMMARY

Knowing the outcomes achieved by health services is essential to being able to achieve the greatest benefit, the best patient care, from the resources used. The effectiveness, efficiency and accountability of the UK NHS all depend on knowing the outcomes it is achieving.

Creating an informed population, able to engage fully in its health care and exercise meaningful choice depends on measures and analyses of outcomes information being available. The groups responsible for commissioning NHS health care for their local populations need information on outcomes to help decide where best to spend their limited funds so as to give the greatest benefit to patients. Information on achieved health outcomes can also help clinicians and managers to compare and improve performance. Measuring health outcomes is a crucial element of assessing NHS productivity; measuring activity without information on patient outcomes is an increasingly unsatisfactory substitute for our real interest: the health benefits achieved for patients.

Many different measures of outcomes are used in research, but none is yet routinely collected and used by the NHS to measure what it is achieving with the huge resources it employs. The time has come to establish outcomes measurement as a matter of routine practice nationally. It was with this strong belief that the Office of Health Economics decided, in the autumn of 2006, to establish a Commission on NHS Outcomes, Performance and Productivity.

This report brings together the results of the OHE Commission's deliberations and the research it commissioned and sets out recommendations for making a reality of outcomes measurement throughout the NHS.

Method

The heart of our approach was evaluation of evidence and discussion of its implications by the members of the OHE Commission. This was taken forward both in face to face meetings and via electronic communication outside those meetings. Between them, the Commission members provide a range of essential skills and perspectives at a senior and expert level: economics, medicine, epidemiology, management and policy; public and private health care sector viewpoints; professional, patient and general public perspectives.

The OHE Commission focussed on four 'tracer' disease areas to cover: acute and long-term conditions, primary and secondary care, life extending and quality of life improving care. The tracer disease areas are: elective surgery; mental health; chronic obstructive pulmonary disease and colorectal cancer. The studies of patient outcome measures in these disease areas all presented evidence on both disease-specific and generic outcome measures, the latter permitting estimates to be made of quality adjusted life years (QALYs) added as a result of health care. There is a tension between the desire to use disease specific measures sensitive to patient health and generic measures that enable health across diseases to be compared. The OHE Commission therefore also investigated whether mapping between condition specific and generic measures is a viable approach.

Although the *raison d'être* of the NHS is to improve health for the UK population, patients have expectations over and above "getting better", most notably about the speed of access and the humanity of the care they receive. Successful health care can also benefit the economy through improved productivity. We commissioned external research to provide further information on these types of benefits: measurement of health service responsiveness; a review of how important different types of benefits have been rated in studies of patient and public preferences; and a review of the literature on indirect cost savings and output increases in the wider economy attributable to health care.

Research Findings – Patients' Health Outcomes

Elective Surgery

As part of its work on health care outputs and productivity the Department of Health commissioned a study of the use of patient reported outcome measures (PROMs) for some common types of elective surgery. In all the procedures the disease specific measure is found to be more responsive than the generic measure of health related quality of life. The report's recommendation is to collect both the generic EQ-5D measure of quality of life and the disease specific measure. There are confounders that need to be adjusted for, so variables such as age, sex, general health status and co-morbidity have to be collected alongside the PROM to be able to compare across centres. BUPA has been collecting outcomes data from patients at its hospitals for nine years, using the SF-36 questionnaire in most cases, but the NHS lacks similar time series of data on a generic or a condition-specific basis.

Mental Health

Mental health differs from the other tracer therapeutic areas in that the link between interventions and outcomes may be less straightforward, and because maintaining health status (preventing further decline) can more often be seen as a positive outcome.

Two mental-health-specific patient outcome measures are options for incorporation into a productivity index. The collection of 'Health of the Nation Outcome Scales' (HoNOS) data, a clinician-led measure, was recommended as part of the Mental Health Minimum Data Set (MHMDS), which was made mandatory by the Department of Health, but coverage hitherto has been poor. The other outcome measure candidate is the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM), a self-rating instrument. No evidence was found to suggest which of these two instruments would be better to use in measuring patient outcomes from mental health services, but we note that CORE-OM has proved popular and workable in community based mental health services and that HoNOS has been recommended for use in more acute phases of illness.

Chronic Obstructive Pulmonary Disease (COPD)

A number of different respiratory-disease-specific measures have been used, of which the Chronic Respiratory Disease Questionnaire (CRQ) and the St George's Respiratory Questionnaire (SGRQ) appear to be best supported methodologically. Alongside collection of disease-specific measures, generic patient outcomes measures such as the EQ-5D or SF- series should be collected.

Colorectal Cancer

Colorectal cancer is one of the commonest groups of cancers and remains a major cause of death. Among the 'tracer' disease areas this is the one in which the ability of treatment to extend length of life is most significant as an indicator of health care outcome.

The research we commissioned revealed a large number of patient outcome measures. The most frequently used instruments were questionnaires developed by the European Organisation for Research and Treatment of Cancer (EORTC) and the Functional Assessment of Cancer Therapy questionnaire (FACT-C). Both were found to be acceptable and feasible to use. In terms of generic measures, both EQ-5D and SF-36 have been found to be valid and reliable in a number of disease areas, with no evidence of inapplicability to colorectal cancer.

Using Disease-Specific Measures to Estimate Generic Measures of Outcome

The extent to which generic measures are correlated with condition-specific measures is variable and not particularly strong. A review of the literature showed that in nearly all cases less than half of the variation in individuals' generic outcome scores could be explained by variations in their measured condition-specific outcome scores. The poor prospects for estimating variations in QALYs, or other generic health outcomes, from disease-specific measures contributes to our view that the preferred way forward should be to collect generic outcomes measures, for example the EQ-5D instrument, alongside disease-specific measures. This appears to be feasible in most disease areas.

Research Findings: Wider Benefits

Health system responsiveness is a concept promoted by the WHO. Responsiveness is defined as the way in which individuals are treated and the environment in which they are treated, and is based on the individual's experience of contact with the health system. The WHO has developed an instrument to measure responsiveness. The same ground is covered by patient experience questions of the National Patient Surveys.

Successful health care interventions may also impact on carers and have benefits to the wider economy. This may have a bearing on the productivity of the NHS and therefore needs to be accounted for. Depression, multiple sclerosis, back pain and stroke have been identified as potentially important disease areas where wider benefits could have a significant impact.

A complete picture of the performance and productivity of health care services requires consideration of wider outcomes concerning the humanity with which health services are provided, and economic impacts outside the NHS. Pragmatism dictates a focused approach, identifying and measuring those wider benefits that are likely to be most significant. Published literature suggests that important benefits to measure include: waiting times, access, care co-ordination, autonomy, choice, communication, confidentiality, dignity, quality of amenities and support for carers.

Implementation Issues

Clinicians

If clinicians do not see the benefits of measuring outcomes in their day-to-day work, then the collection of adequate outcomes data is unlikely to flourish. The rationale for collecting these measures needs to be clearly explained to front line staff, linked to the principles of good clinical practice. Regular and timely feedback of the data (for all stakeholders) is needed to obtain successful implementation. At the level of the individual clinician or clinical team, outcomes data will be valuable for performance management through quality improvement and clinical governance arrangements, revalidation, or to inform patient choice. To sustain the routine use of outcome measures there needs to be training in their use for all clinicians including new doctors.

Patients

Improved patient care is the ultimate reward from improved NHS performance and productivity. In addition, informed patients provide an important lever to stimulate better outcomes from health care, but asking patients to assess what health care has done for them needs to be easy and quick. The use of outcome measures may be particularly useful to patients with chronic conditions as they will be able to track the progression of their disease over time and may consequently be able to take a more active role in their treatment.

Managers

There is management support for the use of outcomes as a means of measuring performance at commissioner level and provider level. Managers are interested in using patient outcomes data to inform the allocative decisions they make and to drive quality improvement of the health care services provided. Training in use of patient outcomes data is likely to be needed. There is an incremental cost involved in collecting patient outcomes data, but the magnitude is modest (£3-£6 per patient). Measurement of patient outcomes and use of that data need to be an explicit element of the government's vision of "world class commissioning".

Infrastructure

IT systems should be clinician and patient friendly to facilitate collection and analysis of outcomes data. But infrastructure does not only apply to IT systems. There is currently a shortage of resources for analysis of the collected data, which needs to be rectified. For data to be fed back appropriately the responsibility for data collection should be formally defined.

Incentives

The collection of outcome measures could be actively encouraged using frameworks already in place. We recommend that initially the emphasis should be on incentivising the collection, analysis and feedback of patient outcome measures, rather than on rewarding the measured levels of outcomes being achieved. In primary care, the Quality and Outcomes Framework (QOF) could be a good tool. In secondary care, Payment by Results (PbR) may be another channel to incentivise uptake by offering premium prices to providers demonstrating the greatest progress in collecting and using outcomes data. Incentives also need to be framed to show that these measures are part of high professional standards through the revalidation process for doctors and the Healthcare Commission core standards.

Nationally promulgated guidelines for patient care may be another avenue for promoting the benefits of outcomes measurement. The Healthcare Commission already runs the National Clinical Audit and Patient Outcomes Programme in the NHS in England, comprising clinical audits in a range of disease areas undertaken in collaboration with medical Royal Colleges and others, and responsibility for which passes to the newly formed Healthcare Quality Improvement Partnership in April 2008. The collection, collation, analysis, feedback and use of patient outcome measures need to be included in Healthcare Commission core standards (Healthcare Commission, 2006). The Audit Commission's proposed assessments of PCTs' use of resources could do likewise (Audit Commission, 2007). Such actions would encourage consistent and standardised measures being applied nationally which allow comparison and would reinforce the seriousness with which implementing outcome measurement is seen.

Phasing and Pace of Implementation

The implementation process is constrained by shortages of skills and resources, but there needs to be a pragmatic implementation timeline. We conclude that while measurement of outcomes can and should be taken forward across a wide range of health services, the speed of implementation should vary. The Department of Health has recently announced, with the publication in December 2007 of its "Standard NHS contract for acute hospital services" (for England) that from April 2009 providers must report patient reported outcome measures for: primary unilateral hip replacement, primary unilateral knee replacement, groin hernia repair and varicose vein procedures. Similarly there is much that could be done to initiate the implementation of patient outcomes measurement in the other condition/treatment areas we have discussed earlier in this report.

Our general approach is to start as early as is practical in 2008/09 with measuring patient outcomes in a small number of diseases and procedures and from 2010/11 to extend to other disease and treatment areas as resources permit. National patient surveys and population sample surveys can be adapted to permit collection of data on the wider impacts of health care and trends over time in economic impacts such as numbers of days lost through illness.

Conclusions and Recommendations

Routine collection and use of outcomes measures in the NHS is both practical and essential. We expect it to lead to improved outcomes, performance and productivity, thereby providing significant benefits to patients. We therefore recommend an early start in a number of major disease areas and on a national scale throughout the UK. In making recommendations, we are seeking a balance between speedy implementation and recognition that it is not practical to apply outcomes measurement everywhere at once. However, it is reasonable to expect that within 5 years outcomes data are collected for the majority of NHS activity, and that within 10 years it would be the exception rather than the rule to find an area of NHS that lacked routine measures of the impact of an intervention on patients in terms of their survival, quality of life and experiences of care. Our recommendations are as follows.

Recommendation

1. Measurement of patient outcomes and use of that information to drive better commissioning of services should be made an explicit element of the Government's vision of "world class commissioning".
2. Mortality and survival data should be collated and published for all condition/treatment areas at the same time as collection of other outcomes data is initiated. This requires data on casemix also to be collected, so that the crude mortality/survival data can be adjusted for the casemix treated. Confidence intervals should be estimated and reported to permit comparisons to be made over time and between providers. The same casemix data can be used to adjust other measures of patient outcomes too.
3. For elective surgery, condition-specific measures should be collected before and after surgery simultaneously with the generic EQ-5D for all NHS patients undergoing common procedures. Implementation should be at the earliest practical opportunity. In its "Standard NHS contract for acute hospital services" published in December 2007, the Department of Health has specified for the first time a requirement for providers to report (from April 2009) patient reported outcome measures for primary unilateral hip replacement, primary unilateral knee replacement, groin hernia repair and varicose vein procedures. We recommend extension to other common elective surgical procedures over the next three years unless emerging experience indicates otherwise.

Action for:

Health departments

Commissioners
Providers
Regulators

Health departments
Commissioners
Providers
Regulators

- | | |
|---|--|
| <p>4. Measurement of mental health outcomes in secondary care should be re-investigated with collection of HoNOS data being incentivised in addition to mental health providers being reminded of its mandatory nature as part of the Mental Health Minimum Data Set. We recommend that, in parallel, routine collection at annual intervals of the CORE-OM measure for all patients receiving community based mental health services be piloted in a range of NHS Trusts at the earliest practical opportunity in 2008/09. If the pilots are successful, collection of CORE-OM data should be rolled out to all providers of community mental health services to NHS patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers
Regulators</p> |
| <p>5. Further research is required to identify a generic measure of health related quality of life that is suitable for application to mental health patients as well as recipients of other kinds of health care. This may be based on an existing measure, such as the Health Utility Index (HUI3) or on adaptation of another measure such as the EQ-5D or one of the SF- series of instruments. The ultimate aim would be to develop a generic measure acceptable in all disease areas, including mental health. Research to that end should be funded by the National Institute for Health Research and its counterparts and commissioned in 2008/09 for completion within a year if possible. Subject to the results of that research, routine measurement of generic patient outcomes (e.g. at annual intervals) alongside the condition-specific measures could then be piloted in 2010/11 in a range of mental health service providers (both hospital and community based).</p> | <p>Health departments
National Institute for Health Research</p> |
| <p>6. The practicality and value of collecting and using measures of outcomes for COPD patients should be piloted by a range of commissioners and their secondary and primary care providers, commencing in 2008/09. The preferred disease specific measure is either the Chronic Respiratory Disease Questionnaire (CRDQ) or the St George's Respiratory Questionnaire (SGRQ). The selected disease specific measure should be collected simultaneously with a generic health related quality of life measure, for which the EQ-5D would be initially acceptable (pending the outcome of the research we propose in recommendation 5), so that outcomes could be compared with those for other conditions/ treatments. If the pilots are successful collection should be rolled out to all commissioners and providers treating and/or managing COPD patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers</p> |
| <p>7. For colorectal cancer patients, mortality data (1 year and 5 years after diagnosis) should continue to be monitored and assessed as an important but partial measure of health service productivity in this disease area given its relatively high rate of mortality. But at the same time we recommend the instigation in 2008/09 of pilot programmes of routine use of the EORTC QLQ-CR38 instrument to measure outcomes in colorectal cancer patients simultaneously with the EQ-5D or an SF- series generic measure. Subject to the results of these pilot projects, collection of QLQ-CR38 and EQ-5D (or an adaptation of it) should be rolled out to all providers of colorectal cancer services to NHS patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers</p> |

- | | |
|--|---|
| <p>8. In addition to patients' health outcomes, we recommend collection of data about the wider benefits from health services by addition of questions on patients' experience of care to the patient questionnaires used to measure health outcomes. Aspects covered should include: access, care co-ordination, autonomy, choice, communication, confidentiality, dignity, quality of amenities and support for carers. Whenever possible this should be done at the same time as the health outcome measures are introduced. Data on waiting times should continue to be used as a measure of one aspect of the humanity of care.</p> | <p>Health departments
Commissioners
Providers
Regulators</p> |
| <p>9. Further research is needed into the relative values of the wider benefits referred to in recommendation 8 and by comparison with health outcomes. This should be commissioned during the course of 2008/9.</p> | <p>Health departments
National Institute for Health Research</p> |
| <p>10. Information on wider benefits is already collected in national patient experience surveys. In the first instance we recommend that, building on its 'customer experience information' project, the Department of Health should work with its counterparts in Northern Ireland, Scotland and Wales on a UK-wide stock-take of patient surveys in process and planned, and then use this as the basis for discussion with all the interested parties how to co-ordinate those surveys in future. The stock-take and discussion should take place in 2008/9.</p> | <p>Health departments</p> |
| <p>11. After a short period of research to identify the most practical way of collecting such data, including working days lost, measurement of productivity effects should be commenced for patients being treated for mental health problems, diseases of the nervous system and diseases of the musculo-skeletal system. Pilot data collection should commence in the year 2009/10. Dependent on the success or otherwise of data collection in these areas, collection of data might in the longer term be spread to all disease areas.</p> | <p>Health departments
National Institute for Health Research
Commissioners
Providers
Regulators</p> |
| <p>12. Modification of existing national surveys of a sample of the general population – such as the Health Survey for England and its Northern Irish, Scottish and Welsh counterparts; the General Household Survey in Great Britain and the Continuous Household Survey in Northern Ireland – should be investigated as one way of collecting at a general level data relevant to the wider societal impacts of health care interventions, including impacts on carers.</p> | <p>Health departments</p> |
| <p>13. Health care professional bodies, including the Royal Colleges, should explicitly and actively support the collection and use by their members of patient outcome measures, including the provision of training in their collection and use for both existing and new clinicians.</p> | <p>Health care professional bodies
Medical schools</p> |

- | | |
|--|--|
| <p>14. Training in understanding and using outcome measures should be provided to NHS commissioner and provider managers.</p> | <p>Health departments</p> |
| <p>15. Investment should be made in IT and appropriately trained staff to enable user-friendly information collection, collation and analysis of outcome measures.</p> | <p>Health departments
Commissioners
Providers</p> |
| <p>16. The NHS should, commencing as early as possible in 2008/09, pilot, monitor and evaluate explicit incentives to collect, collate, analyse, feedback and use patient outcome measures. Initially at least this would not be extended to rewarding providers for the measured levels of outcomes achieved, although that could be considered once experience with use of outcomes measures has accumulated. In primary care, the Quality and Outcomes Framework (QOF) of the GP contract could be a good tool, i.e. GP practices could be paid for achieving high coverage in administering patient outcomes questionnaires that enable collection of outcomes data. Other commissioners and providers might be incentivised financially through parallel arrangements. Publishing the outcomes collected, by organisation, will also be a vital part of the incentivising process. Good commissioners and providers will want to be seen to provide that information to the populations they serve.</p> | <p>Health departments
Commissioners
Providers
Regulators</p> |
| <p>17. There should be an expectation that within 5 years (i.e. by the end of 2012/13) routine measures of patient outcomes comprising the impact of an NHS intervention in terms of patient survival, quality of life, and experience of care, are collected for the majority of NHS activity. Within 10 years it should be the exception rather than the rule to find an area of NHS activity that is not assessed and actively managed according to the outcomes achieved.</p> | <p>Health departments
Commissioners
Providers
Regulators</p> |
| <p>18. The NHS should work with the ONS Centre for the Measurement of Government Activity to ensure that the health outcomes data used by the NHS can also be used by ONS in its periodic assessments of NHS productivity and, subsequently, in the National Accounts.</p> | <p>Health departments
ONS</p> |

1. INTRODUCTION

1.1 Rationale and Terms of Reference

It is widely understood that providing health services uses up a lot of resources: about one pound in every 11 spent in the UK goes on health care, and most of that is spent on NHS care. Large quantities of data are available on the inputs used by the NHS and expenditure on them – staff, facilities, equipment – and on the activities undertaken, such as consultations with GPs or spells of hospital care. But we know much less about the outcomes those resources and activities produce.

A large range of measures of the outcomes of health care are in widespread use in research projects and clinical trials. Many of these measures are disease/condition specific but a number of more generic measures of health outcome are also gaining currency in research. None of these measures, however, is yet being collected routinely outside of specific research projects.

Knowing the outcomes achieved by health services is essential to being able to achieve the greatest benefit, the best patient care, from the resources used. The effectiveness, efficiency and accountability of the NHS all depend on knowing the outcomes it is achieving. An informed population, able to engage fully in its health care and exercise meaningful choice, depends on measures and analyses of outcomes information being available. The groups responsible for commissioning NHS health care for their local populations need information on outcomes to help decide where best to spend their limited funds so as to give the greatest benefit to patients. Information on achieved health outcomes can also help clinicians and managers to compare and improve performance. Measuring health outcomes is a crucial element of assessing NHS productivity; measuring activity without information on patient outcomes is an increasingly unsatisfactory substitute for our real interest: the health benefits achieved for patients.

Much work is already being done to advance that knowledge but the time has come to establish outcomes measurement as a matter of routine practice nationally and not just a subject of research and local experiment. It was with this strong belief that the Office of Health Economics decided, in the autumn of 2006, to establish a Commission on NHS Outcomes, Performance and Productivity with the terms of reference:

- To investigate what patient outcome data the NHS could and should practically collect on a routine basis to secure better measurement of NHS productivity and therefore to improve performance.
- To make explicit recommendations on analyses and reporting arrangements for such data to ensure that the outcome/productivity measures can be utilised by NHS decision-makers, especially commissioners, to inform allocative efficiency issues at both local and national level.
- Identifying measures that are sufficiently generic that they can be aggregated and used at a national level.

The OHE Commission was chaired by Professor Peter Smith, Director of the Centre for Health Economics at the University of York. Other members are eminent experts:

Nick Black	Professor of Health Services Research, London School of Hygiene and Tropical Medicine
Roger Boyle	National Director for Heart Disease, Department of Health
Angela Coulter	Chief Executive, Picker Institute
Nancy Devlin	Professor of Health Economics and Head of Economics Department, City University
Nigel Edwards	Director of Policy, NHS Confederation
Mike Richards	National Cancer Director, Department of Health
Adrian Towse	Director, Office of Health Economics
Andrew Vallance-Owen	Group Medical Director, BUPA

This report brings together the results of the OHE Commission's deliberations and the research it commissioned. It is our hope that the report will be useful for a wide variety of audiences who will have a number of different perspectives, but who share the ultimate objective of improving patient care. At the local level, providers of health care, and the clinicians who deliver that care, need information on outcomes to improve the quality of care. Patients will require performance measures to enable them to compare different providers. The commissioners of care – primary care organisations and GP practices – need information on performance and productivity to make decisions on the allocation of scarce health care resources, and at the National Accounts level there is need to assess the change in productivity of the NHS given the resources being invested.

The essence of the work and recommendations presented in this report may perhaps be captured in the prosaic words: "routine" and "pragmatic". Our aim has been to identify what may be done in the NHS on a routine basis to enable assessment of the outcomes of health care services. Routine implementation requires that recommendations about the measurement, collation and analysis of outcomes data be pragmatic.

We note with great pleasure that the 'Operating Framework for the NHS in England 2008/09', which was published while we were drafting our report, makes a breakthrough in this direction by requiring that NHS providers of elective hip and knee replacements, groin hernia surgery and varicose vein procedures collect and report from April 2009 patient reported outcome measures (PROMs) for that activity (Department of Health, 2007c).

1.2 Levels at which Outcome Measures are Needed

The OHE Commission focuses on three separate levels of the NHS, as follows, but in all cases on the basis of two underlying principles, which are that, to be most useful, the measurement of outcomes needs to produce data that:

- are comparable across the NHS, which implies the need for a consistent approach to be adopted nationally; and
- apply to care however and wherever it is provided: as part of primary, secondary or tertiary health care services; in patients' homes, GPs' surgeries, hospitals or any other setting.

Provider level

Providers and the individual clinical teams within them can use condition-specific outcome measures to understand how well they are performing relative to their peers and how their performance is changing over time. A provider of cancer services, for example, can compare their cancer-specific outcome measures with those achieved by other providers. Relating provider-specific, disease-specific outcomes to provider-specific costs in that disease area of care would permit efficiency comparisons between providers at the level of a specific service.

But any single NHS Trust or other provider organisation will be delivering a mix of different services to patients with different kinds of diseases requiring different treatments. Many of the costs a provider incurs are common: that is, some costs would still be incurred even if the mix of services provided were to change. For example, substantial elements of the costs of theatre time (although not of the specialised surgeons) are common to the delivery of heart surgery and hip surgery. Hence, allocation of common costs between different services inevitably contains an element of arbitrariness. Comparing the overall efficiency of providers delivering different mixes of services becomes feasible if the outcomes of different services can be aggregated. To do that requires generic – not disease-specific – measures of patient outcomes.

Commissioner level and NHS overall

Disease specific outcome measures can help commissioners of health care to understand how well providers are performing relative to one another and relative to (efficient) expectations.

However, decisions about how much to spend from a limited commissioning budget on one type of services versus all others across all disease areas require information on the value gained and lost as resources are switched from one treatment or group of patients to another. This requires a generic outcome measure that applies across all disease groups so that comparisons can be made. The results can then be analysed in terms of a marginal cost per QALY (quality adjusted life year) league table, or a monetary value could be given to outcomes.

Resource allocation within the NHS, i.e. allocation between groups of patients with different diseases, requires that outcome measures at some point become generic. However, the OHE Commission has also explored the short/medium term advantages of using existing disease specific outcome data if they are available, until such time as data concerning generic measures become widely available.

Furthermore, many performance management tasks – e.g. reducing differences in efficiency or productivity between providers of similar treatments to comparable groups of patients – do not require generic outcome measures. Comparisons within disease areas and over time can be extremely helpful for performance management.

National Accounts level

Measuring the aggregate productivity and the performance of the NHS is complex. We define productivity as the ratio of a measure of total outputs to a measure of total inputs. Productivity change in a time period measures whether the proportionate change in outputs is greater than the proportionate change in inputs. At present NHS output is measured in terms of activities covering around 80% of spend in England. Data are obtained from a number of different sources. The main ones are: NHS Reference Costs and Hospital Episode Statistics for hospital procedures, the General Household Survey for primary care consultations, and the NHS Prescription Pricing Division for the

numbers of prescriptions. Other measures capture the amount of use of NHS Direct and of dental services, and the numbers of sight tests and of ambulance emergency journeys. Overall NHS activity and growth is then measured by a cost-weighted activity index.

As the primary aim of the NHS is to improve and maintain the health of the population, it is important to have some assessment of its ability to do this in any measure of productivity. We want to measure how NHS outputs (as measured by activity) improve population health outcomes. We need to adjust NHS output to take account of the quality of the health outcomes it achieves. Health can be influenced by many lifestyle and related factors other than NHS health care, such as smoking and diet, so outcomes attributed to the NHS need to reflect the contribution the NHS is making to improving health. People also have expectations about how NHS care is delivered. The quality of delivery is also an important element of output. Aggregate National Accounts measures of NHS output and productivity need therefore to adjust activity based measures of output to take account of the delivery of improved patient quality and quantity of life if they are to reflect the underlying performance of the NHS.

In the remainder of this section of the report we briefly set out summaries of recent pieces of work that have helped to move the debate on measuring productivity forwards and which provide important context for the work of the OHE Commission.

1.3 Context

National Centre for Health Outcomes Development (NCHOD)

The Department of Health initiated a national approach to health outcomes development in 1993 by setting up within it the Central Health Outcomes Unit. The Unit of Health Care Epidemiology (UHCE) at the University of Oxford was first involved in 1996 when ten condition-specific studies (including asthma, MI and diabetes) to develop menus of outcome measures were commissioned. In 1998 the Department of Health decided to outsource the function and the NCHOD was established. NCHOD comprises teams from UHCE and the London School of Hygiene and Tropical Medicine (LSHTM). The research and development work is led by UHCE, whereas LSHTM is responsible for the production and publication of outcome measures. NCHOD publishes condition-specific outcome indicators through its compendium of clinical and health indicators, available on their website and updated annually. However this compendium does not contain any patient assessed health indicators and contains no quality of life data or other outcome measures linked to the receipt of health care interventions.

NCHOD also undertakes work in population-based outcome measures, patient-reported health instruments, and outcome indicators derived from linked Hospital Episode Statistics and ONS mortality data. The patient-reported health instrument section is a searchable database which includes systematic reviews of patient-reported outcome measures.

Measuring success in the NHS – King's Fund, funded by Dr. Foster (2004)

The King's Fund report (Appleby and Devlin, 2004) highlighted the lack of routinely measure health-related quality-of-life (HRQoL) in the UK. There is much to be gained from routine before and after measurements of health state both as a means of monitoring patient health and to measure the performance of providers. In summary, the authors considered that HRQoL data have the potential to strengthen the management of clinicians, surgical teams and hospitals, by shifting the emphasis from process, activity and intermediate outputs to outcome. However, they also concluded that there needs to be more information before implementation on: costs and benefits; which measures

discriminate best between good and bad providers; and how frequently these data need to be collected.

Quest for Quality and Improved Performance (QQUIP) – research initiative of The Health Foundation (2005 onwards)

This five-year initiative commenced in 2005. The aim is to collect a wide range of data about current performance and capacity for improvement in terms of quality and cost-effectiveness – looking at where health care resources are currently being spent and where the best value for money is provided. QQUIP intends to provide data for decision-makers within health care systems. It is split into three workstreams: quality information centre, value for money analyses, and quality enhancing innovations:

- The quality information centre is a database that collates all available data on quality and performance.
- The value for money stream is to look at how NHS outputs are converted to outcomes, and estimate costs and benefits in a common scale so that disease areas can be compared. This may provide a way of comparing costs and benefits of different national policies.
- The quality enhancing interventions stream includes evidence based reviews of the effectiveness of interventions and therefore their ability to increase the quality of health care.

Public disclosure of cardiac surgery results in the UK NHS (2006)

The public disclosure of cardiac surgery results was a response to the report of the public enquiry into the events at Bristol Royal Infirmary published in 2001, to increase the public accountability of the cardiac surgeons. In 2005 cardiac surgery mortality data were made available to named cardiac surgeon level in The Guardian newspaper through a request under the Freedom of Information Act. This was followed in 2006 by cardiac surgery mortality data being published on a website run by the Healthcare Commission and the Society of Cardiothoracic Surgeons of GB and Ireland (<http://heartsurgery.healthcarecommission.org.uk>). The website displays data on rates of survival for particular surgeons and hospitals, collected from one of the key national audits which the Healthcare Commission oversee. The rates of survival are risk adjusted taking into account 15 potential factors.

Supporters of public disclosure expect that such an exercise will lead to quality improvement as patients and commissioners may select units with the best results, and therefore motivate hospitals to improve. However others feel that surgeons may display risk averse behaviour when deciding to operate on a particular patient, which is supported by a number of surveys of attitudes of clinicians.

A study was published early in 2007 which investigated the effect of public disclosure of cardiac surgery results in the UK (Bridgewater et al, 2007). There was a significant decrease in crude and risk-adjusted mortality after the introduction of the public disclosure, without any evidence of high risk patients being denied operations – there was no decrease in the number of high risk cases coming to surgery. The cardiac surgery experience therefore gives a practical example of how publication of outcomes may lead to quality improvement.

Patient reported outcome measures (PROMs) in elective surgery – LSHTM/RCS report to the Department of Health, December 2007

The Department of Health commissioned, in 2005, a study of the use of patient reported outcome measures (PROMs) in elective surgery. The study was undertaken by the Health Services Research Unit of the London School of Hygiene and Tropical Medicine (LSHTM) and the Clinical Effectiveness Unit of the Royal College of Surgeons of England (RCS). Its report to the Department of Health was published in December 2007 (Browne et al., 2007).

The research team reviewed the PROMs used in five elective surgical procedures and recommended a disease-specific measure and a generic measure for each procedure based on the measures' psychometric properties as well as their practicalities. The measures were then piloted to explore the feasibility and practicalities of data collection, analysis and presentation. As a result, the LSHTM/RCS team recommended the following disease-specific measures for collection in the four surgical procedures:

- Hernia repair: none
- Hip surgery: Oxford Hip Score
- Knee surgery: Oxford Knee Score
- Varicose vein surgery: Aberdeen Varicose Vein Questionnaire

(Concerns about the validity of the measures for cataract surgery meant no recommendation was made.)

Alongside these measures, the research team recommended that the EQ-5D instrument be collected as the generic measure of health gain. The average cost per matching pre-operative and follow-up questionnaire collected was found to be in the range £5.85-£6.58 (depending on type of surgery).

More information about the results of this study is in Section 3.1 below.

Following the LSHTM/RCS work, the Department of Health's "Operating Framework for the NHS in England 2008/09", published in December 2007, requires for the first time that NHS providers of elective hip and knee replacements, groin hernia surgery and varicose vein procedures must from April 2009 collect and report patient reported outcome measures (PROMs) for that activity (Department of Health, 2007c).

Atkinson Review

The review on measuring government output and productivity across all areas of public services, including health care, was commissioned in December 2003 and reported in January 2005 (Atkinson, 2005).

For the purposes of the Atkinson Review, "output" in the context of National Accounts means the contribution of a sector to the UK's Gross Domestic Product (GDP). GDP can be estimated via three different routes, which in principle should yield the same answer, by measuring respectively: income received by all UK residents; expenditure in the UK economy (including exports less imports); or outputs of UK businesses and of the UK public sector. Measuring changes in output at this aggregate level are equivalent to measuring changes in national income. Therefore measurement of the changing quality as well as the changing quantity of public service non-marketed output was a central theme of the report.

It is important to note that because the Atkinson Review relates to national income accounting approaches it reported within the framework of EU and OECD conventions for National Accounts reporting which require a focus on outputs rather than on outcomes. Hence the Atkinson Review argues for outcomes to be used to quality adjust activity-based measures of output. For the NHS itself it is possible to focus directly on outcomes as well as using information on outcomes to quality adjust measures of activity.

The Atkinson Review made 54 recommendations. Its general recommendations included the:

- continued use and development of direct output measures for public services;
- adjustment of measures of output growth to take account of quality change, recognising that "the greater degree of subjectivity in making quality adjustments, compared with volume measures, means a higher acceptability threshold should be set for their introduction into the National Accounts;
- use of a process of "triangulation" to explore productivity change in the context of "a range of other information" in recognition that no one number "can fully capture the performance of complex public services with multiple objectives."

It had three particular areas of focus in health:

- better measures for output in primary care;
- movement towards measuring whole courses of treatment;
- ideas for measuring quality change in health care, including "options for collecting new information on health outcomes resulting from NHS treatments".

In the context of health care, the report also recognised that cost-based weights and counting activities ignoring the quality of activities in terms of their outcomes could actually be perverse in areas such as public health interventions, which reduce the need for treatment. Increased spend when successful leads to less NHS activity overall.

Since the publication of the Atkinson Review final report, there have been several further publications that have advanced thinking in measuring the quality of NHS output. The ONS established the UK Centre for the Measurement of Government Activity which has published and sponsored a number of papers looking at health care productivity. We discuss this work in our Chapter 5 discussion on aggregation issues.

York/NIESR Report 'Developing new approaches to measuring NHS outputs and productivity'

This important study was commissioned by the Department of Health to run in parallel to the Atkinson Review. It was undertaken jointly by the University of York and the National Institute for Economic and Social Research (NIESR).

The York/NIESR research project (Dawson et al., 2005) looked at a number of interacting factors in seeking to develop a quality adjustment for NHS output. The major ones they sought to incorporate in an index were as follows:

- Improving survival rates so reducing avoidable deaths. The York/NIESR project used data on death within 30 days of hospital admission produced by NCHOD and adjustment was made for the fact that some conditions have a higher death rate.

- Health effects. Ideally data on health states of patients should be collected before and at intervals after all treatments so as to provide a dataset from which to measure health effects, but this is not yet done. In the York/NIESR paper illustrative data were used for 29 procedures using two sources: data from BUPA and data from other research studies. There are limitations with these data, and wider coverage of health effects would be more advantageous. There are also difficulties in making extrapolations to health gain over patients' lifetimes given the short term nature of the measures used.
- Adjustments for life expectancy, reflecting that older people have less time to benefit from treatment, whilst acknowledging that it would be preferable to use condition-specific and age-specific survival rates.
- Waiting times for treatment were also taken into account on the grounds that longer waits may reduce the health gain from an intervention and make the intervention more stressful for the patient.

Department of Health "Healthcare output and productivity: accounting for quality change"

This paper by the Department of Health (2005a) responded to the York/NIESR project and to the Atkinson Review. In it a number of other factors additional to those identified by the York/NIESR Report were put forward for inclusion in a quality adjustment. The main ones were as follows:

- There are a number of areas where the cost weights and marginal valuation may be very different, which may therefore cause a difference in measured NHS output. As there is much data on coronary heart disease (CHD), the Department of Health explored the example of statins use and showed that using value weights instead of cost weights in the index for outputs could increase recorded annual NHS output growth by 0.81%, because of the added life years as a result of statin therapy.
- In addition, the Department of Health explored whether there has been an improvement in primary medical care using data from QRESEARCH (a large general practice database). In the Department of Health paper, data showing improved hypertension and cholesterol control were used for quality adjustments.
- Other quality adjustments that the Department of Health investigated included longer survival rates from myocardial infarction (MI), which they attributed to improvement in health care particularly following the CHD National Service Framework (NSF). There may be further work to see if this adjustment can be applied to other parts of the CHD NSF. In addition, adjustments for patient experience were explored. NHS organisations carry out a number of surveys of patients' experience of health care services, but the impact of incorporating these responses into quality adjustments was small.

Evidence on quality improvement from the US

The rapid growth in US health care expenditure and of the prices of drugs, consultations and other interventions prompted academic research in 1990s in the US to quality-adjust prices and outputs to understand the extent to which indices of price inflation and expenditure growth can be adjusted to take account of better quality outcomes for patients. A number of US studies explored direct measurement approaches, advocating focus on particular diseases, with treatment costs and health outcomes being estimated for that disease. Economists in the US published numerous articles based on this methodology. Overviews of early work are given in Triplett (1999) and Cutler and Berndt

(2002). The following paragraphs summarise some of the findings in the two main disease areas examined – cardiovascular disease and mental health.

Cutler and McClellan (2001) looked at the evidence on productivity change in the treatment of heart attacks. They looked at survival rates for fee-for-service Medicare patients who had a heart attack between the 15 year period 1984 and 1998. Treatment options had greatly expanded in this area both in terms of medical therapy and surgery, namely thrombolytics and coronary artery bypass grafts respectively. The costs of the treatment of heart attacks in US increased from \$3 billion in 1984 to \$5 billion in 1998 which was a real terms growth of 3.4% annually. The incidence of heart attacks reduced by 0.8% per year, due partly to better risk assessment and prevention through smoking policies, reduction of blood pressure and cholesterol. Spending per case therefore increased by 4.2%, increasing costs per case by \$10,000. Both quality of life and longevity improved, but there was only good data available on the latter. Life expectancy increased from five years to six years over the 15 year period. However, there is an issue of attribution. Not all of the one year addition to life expectancy is a result of the extra \$10,000 cost of improved treatment.

In related studies Cutler et al (1998, 2001) developed a cost-of-living index for cardiovascular treatment designed to reflect improvements in health outcomes. In the 2001 paper they calculated that the quality adjusted cost of living index for heart attack treatment has fallen by 1.7% per annum over the period 1984-1994. Cutler et al (2006) evaluated cardiovascular disease disability reductions from intensive medical care, and showed a 50-70% reduction in disability using appropriate pharmaceuticals and invasive procedures, which equated to an addition of 3.7 years of quality-adjusted life expectancy. Most recently, Cutler (2007) estimated, using 17 years of follow up heart attack data, that revascularisation was associated with a year of additional life at a cost of about \$40,000.

Berndt and colleagues (Berndt et al 2001, 2002) carried out a number of studies in the area of depression, following the launch of selective serotonin reuptake inhibitors (SSRIs) to treat depression in the early 1990s. Evidence showed that SSRIs were of similar cost to the existing tricyclic anti-depressant medicines (TCAs), but had better health outcomes; i.e. the efficacy was equivalent but SSRIs had a lower risk of overdose. Extra treatment costs were \$1,000 per case, but the quality of life benefits were equivalent to \$3,000. Berndt et al (2002) estimated that the quality adjusted cost of successfully treating an episode of acute phase major depression fell over the period 1991-6 by between 1.7% and 2.1% per annum.

In a similar exercise Berndt and colleagues looked at schizophrenia (Frank et al, 2004) using Medicaid data over the period 1994-5 to 1999-2000 and found that whilst the annualized costs for the ongoing treatment of schizophrenia per person had increased at about 0.5% per year, when adjustments were made for patient mix and the quality of treatment outcomes mean treatment costs had fallen about 5.5% per year. Berndt et al (2005) extended the analysis to bipolar disorders and found that changes in treatment mix and the quality of outcomes for the period 1991-5 time period equated to a mean rate of price decline of 3.3% per annum.

Measuring health care output in the UK: a diagnosis based approach (2004)

Using the US literature on price indices developed by Cutler and others, Nicola Mai, a member of the Atkinson Review team, proposed a method of 'calculating health care output in the presence of innovation and treatment substitution'. The *cost weighted activity index*, is adapted for disease and patient to give two different indices. The *cost weighted patient index* reflects the average cost not just of a single activity but of the whole treatment pathway a patient is taken through, hence it introduces links between activities through treatments. The *cost weighted disease index* accounts for the variety of treatments that can be used in a particular disease – hence is sensitive to substitution

of treatments. In other words, substitution of a lower cost treatment that is equally effective does not lead to a reduction in output as occurs with the current cost weighted activity approach.

Mai gives the example of treatment of coronary disease to test for the sensitivity of all three indices and highlights the advantage of the cost weighted disease index. This approach shows the advantages of a disease based rather than intervention specific approach.

1.4 Structure of the Report

Thus, from a variety of perspectives, assessing the outcomes, performance and productivity of health care services has been the subject of increasing attention. The OHE Commission intends to take matters a stride further by recommending pragmatic ways in which measurement and use of patient outcomes can routinely be implemented in the NHS, starting now.

The remainder of our report is structured as follows. The way in which we did our work and commissioned research inputs is described in Section 2. The findings from the work undertaken are presented in two main sections. Findings with respect to patients' health outcomes are brought together and discussed in Section 3, while Section 4 covers findings concerned with other possible benefits of health services. Issues raised by the need to aggregate outcomes across the range of NHS services for national accounting purposes are discussed in Section 5. The way in which measurement of outcomes is implemented will be crucial to their success; Section 6 covers the major implementation issues. Conclusions and recommendations are then brought together in Section 7.

2. METHOD

2.1 The Approach Taken by the OHE Commission

The method of working adopted by the OHE Commission was designed to maximise the effectiveness of its members' inputs and to come up in a little over one year with practical ways forward that would be of value to the NHS. Between them, the Commission members provided a range of essential core skills and perspectives at a senior and expert level: economics, medicine, epidemiology, management and policy; public and private health care sector viewpoints; professional, patient and general public perspectives.

The heart of our approach was focused on consideration of evidence and discussion of its implications by the members of the OHE Commission. This was taken forward both in face to face meetings – five of them between October 2006 and November 2007 – and via electronic communication outside those meetings. The Commission's secretariat provided briefing material for the meetings and discussions, and followed up the actions they produced.

The starting point of the OHE Commission's work was a review of the recent literature on measuring NHS productivity, much of which is discussed in the preceding chapter. In addition to the survey of existing literature and ongoing work undertaken by the Secretariat, and to the input of the Commission members themselves, funding was raised from external sources and used to pay for research projects on topics identified by the Commission in its discussions. The researchers were asked to review published literature and data sources, consult experts in the respective fields, and provide written reports to the OHE Commission. These research reports are available on the OHE Commission website www.ohe.org/page/ohecommission. The final major source of inputs was obtained from two workshops organised by the Commission secretariat. Figure 2.1 illustrates the main elements of the work process.

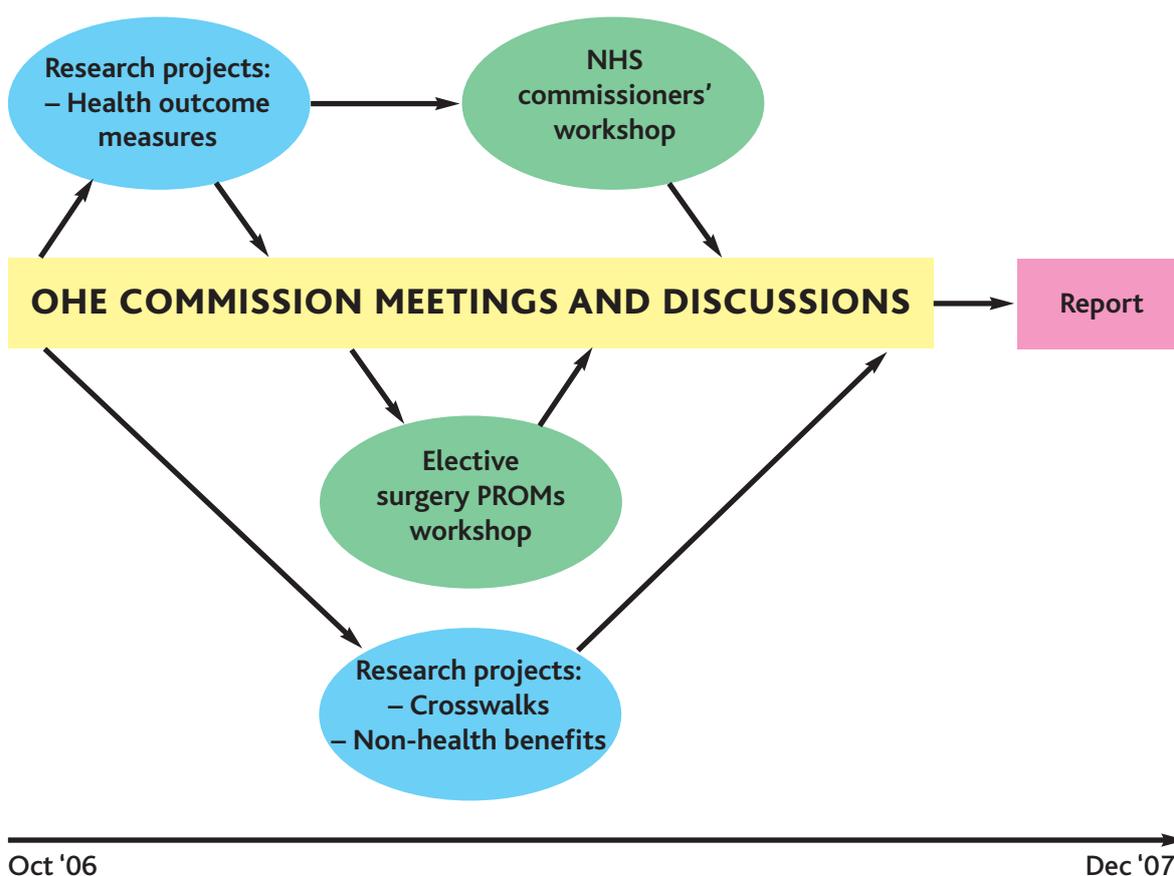
At its first meeting the OHE Commission decided that a focus on identifying the potential benefits of tracking patient outcomes in a small number of 'tracer' disease areas would be more beneficial than attempting to leap immediately to identifying global measures of health service productivity. The 'micro' measures found at the disease level might then be aggregated to provide 'macro' measures of NHS productivity. Such macro measures could in turn be used by ONS to assist in its determination of public sector productivity as part of national income accounting.

In order to be able to test out the issues raised by outcome measurement across the spectrum of NHS activities and to draw broad conclusions, the OHE Commission concluded that a mix of 'tracer' disease areas was needed so as to cover: acute and long-term conditions, primary and secondary care, life extending and quality of life improving care. The following four disease areas were selected for investigation so as to cover these dimensions:

- Elective surgery – for hip and knee replacement, hernias, varicose veins, cataracts
- Mental health
- Chronic obstructive pulmonary disease (COPD)
- Colorectal cancer

The evidence on the likely value of routine use of patient outcome measures in these four areas is summarised in Section 3 of this report.

Figure 2.1: Main elements of the OHE Commission's work process



2.2 The Research Commissioned and Workshops Held

For elective surgery we were able to draw on the research commissioned by the Department of Health from the London School of Hygiene and Tropical Medicine (LSHTM) to study routine use of patient reported outcome measures (PROMs) in short-stay elective care settings for high-volume routine surgery, covering cataract surgery, hip and knee replacement, varicose vein procedures and hernia repair (Department of Health, 2005b). A member of the OHE Commission, Nick Black, was a leading member of the LSHTM research team for this project and was able to brief the Commission on the evidence published in this area (including Smith et al., 2005).

Some of the findings and implications of the LSHTM research were discussed at a workshop with other organisations with practical experience of measuring PROMs for elective surgery, namely BUPA and CHKS. The workshop was hosted by the OHE on 16th July 2007. Speakers from LSHTM (Nick Black), BUPA and CHKS presented and compared information from their respective work on using PROMs to routinely measure health care outcomes from common elective surgical interventions. The information presented was discussed and challenged by an invited audience drawn from the NHS, academia and policy makers. A summary of the workshop results is presented at Appendix 10.

For the other three 'tracer' areas listed above we commissioned specific research to address five groups of questions:

1. What outcome measures are currently used within the disease area? Which of these are believed to be the most sound methodologically and which are used most commonly in practice?

2. What are the practical issues about collecting these outcome measures routinely? How difficult would it be to collect these measures/data within the NHS? Are clinicians likely to use them?
3. Can these disease specific outcome measures be converted into a QALY or some other generic measure? How easily can the disease specific measures be mapped onto the generic measure? Could these outcome measures be used to monitor changes in productivity?
4. Is there a time series of these outcome measures available; going back how many years? If not, were there measures used previously in this disease area and could these be mapped to the current outcome measure to construct a time series?
5. Are any data available on changes in the mix of treatments in this disease area over time? Are there data relating to the costs of the changing mix of treatment?

The reviews of outcome measures were undertaken by:

- mental health – Rowena Jacobs of the University of York Centre for Health Economics;
- COPD – Helen Starkie and Andrew Briggs of the University of Glasgow, Department of Public Health and Health Policy;
- colorectal cancer – Paul Trueman and colleagues at the University of York, York Health Economics Consortium.

Summaries of the three reports are attached at Appendices 2-4 and the full reports are available on the OHE Commission website at www.ohe.org/page/ohecommission.

The researchers presented their findings on the answers to the five groups of questions to an audience drawn from the NHS, academia and policy circles at a second workshop organised by the Commission Secretariat and hosted by the OHE on 12th October 2007. The practicality and usefulness of the different types of outcome measures identified by the researchers were tested and challenged by speakers representing different elements of NHS commissioning of health care: practice based commissioning GPs, primary care organisations, strategic health authorities and the Department of Health. The findings from the October workshop are summarised at Appendix 11.

The studies of patient outcome measures used in health services concerned with elective surgery, mental health, colorectal cancer and COPD all presented evidence on both disease-specific and generic outcome measures. For many purposes disease specific measures may be appropriate but whenever outcomes from services in different disease areas are to be compared or aggregated there is a need for a common currency to be applied in all areas, i.e. some kind of generic outcome measure(s). Sometimes in these areas generic measures have been collected in addition to disease specific measures, but sometimes not. This is typical of many disease areas, where there is a tension between the desire to use disease specific measures that are sensitive to patient health and more generic measures that enable health across diseases to be compared.

The OHE Commission therefore asked John Brazier (Professor of Health Economics at the University of Sheffield School of Health and Related Research) and colleagues to investigate whether mapping between condition specific and generic measures is a viable approach. That is, where disease specific outcome measures have been collected but not generic measures, can estimates of the latter be obtained from the former, under what circumstances, and how? Professor Brazier and colleagues' findings are discussed in Section 3 below and summarised in Appendix 8. Their full report is available on the OHE Commission website at www.ohe.org/page/ohecommission.

Although the *raison d'être* of the NHS is to deliver improved health for the UK population, there are other impacts of the services it provides. Patients have expectations over and above "getting better", most notably about the humanity with which care is provided. The quality of patients' experience when receiving NHS care can be better or worse on numerous dimensions in addition to the health gains obtained from that care. These dimensions include:

- speed of access to advice or treatment;
- participation in decisions;
- respect accorded and dignity preserved;
- availability of comprehensible information about treatment including provision of support for self-care;
- attention to physical and environmental needs.

Furthermore, successful health care can also benefit the economy through improved productivity: it may mean less time spent off work by patients and/or their carers, and greater productivity when at work.

We commissioned three pieces of external research to provide further information on these types of benefits and how they might be taken into account when measuring the total outcomes generated by NHS activity:

- Measurement of health service responsiveness. In 2000 the World Health Organisation (WHO) published a controversial report in which it attempted to determine the relative success of different countries' health care systems (WHO, 2000). One major element of this was assessment of health services' responsiveness to patients' needs, what we refer to as the humanity of care. We commissioned Nigel Rice and Silvana Robone of the University of York Centre for Health Economics to report on how measurement of responsiveness has been developed since 2000 and issues arising from using such measures in practice.
- A review of how aspects of patient experiences and health benefits have been evaluated in studies assessing health services and technologies. We commissioned Emma McIntosh of the University of Oxford Health Economics Research Centre to carry out a systematic review of studies using the technique of discrete choice experiments (DCEs) in health care to find which attributes have been tested and how much, relative to health gains, they have been found to matter to patients. DCE studies require patients to specify the relative value of different aspects of care, so it is possible to understand not just "what matters" but how much it matters to the patient in their overall experience of health care.
- A review of the literature on indirect cost savings and output increases in the wider economy attributable to health care. This was undertaken by Clive Pritchard and Albert Chibi of the Office of Health Economics.

These three pieces of work and their implications for routine measurement of NHS productivity are discussed in Section 4 of the report.

In addition to the research projects and workshops so far outlined, the OHE Commission benefited from informal discussions with external advisers of issues concerning the aggregation of varied micro-level measure of outcomes into aggregate measures suitable for assessing the productivity of the NHS as a whole. The Commission is particular grateful to Diane Dawson (co-author of the

York/NIESR study for the Department of Health, Dawson et al., 2005) and David Parkin (Professor of Health Economics, City University) for their challenging and thought-provoking inputs. In addition to his informal input, the OHE Commission asked Professor Parkin to report on:

- the crucial issue as to whether condition specific measures of health, generic health status measures, the effects of health on workforce productivity, and the humanity of care (satisfaction, reduced waiting times, responsiveness to preferences with respect to process of care, and so on), can be brought together in a way that yields robust evidence about changes in NHS productivity;
- whether such a measure of NHS productivity can be constructed in a way that is both helpful to the NHS at the national level in understanding overall NHS performance trends but which is also consistent with the national accounting framework. The latter is important if national accounts are to better reflect public service productivity by adjusting for improvements in the quality of activities, i.e. by looking at changes in the outcomes they achieve over time.

This piece of work and its implications for measuring aggregate NHS performance and for National Accounting is discussed in Section 5 of the report.

3. RESEARCH FINDINGS – PATIENTS’ HEALTH OUTCOMES

As stated in the introduction, we commissioned research to look at patient outcome measures within four ‘tracer’ disease areas: elective surgery, mental health, COPD and colorectal cancer. Thus we cover both acute and chronic illnesses; health services aimed largely at improving survival, i.e. quantity of life, and others aimed largely at improving quality of life; and we cover both primary care and secondary care. Summaries of the reports of these research pieces are provided in appendices to this report, and the full reports may be read and downloaded from the OHE Commission website (www.ohe.org/page/ohecommission).

At a basic level it is important to collect treatment-specific and (where possible) provider-specific data on survival rates at different intervals after initiation of treatment. For the majority of treatments, where mortality is very low, this may not reveal much about the quality of outcomes being achieved in general, but it is nevertheless an important component of the overall outcomes picture. But for treatments such as for some cancers (see Section 3.4) mortality rates are still relatively high and a commensurately bigger part of the outcomes story.

In each disease area a number of different measures of patient outcomes, specific to the condition and generic, are available. The standard psychometric criteria for assessing the relative merits of different outcomes measures are:

- Reliability – concerned with accuracy – can results be reproduced consistently over time?
- Validity – does the instrument measure what it attempts to measure in the different settings it is applied to?
- Responsiveness – is it sensitive to clinical change?
- Precision – can it clearly distinguish between respondents?
- Practicality – is its use acceptable and feasible?

In the following pages we take each therapeutic area in turn, highlight the most important findings and then draw out our general conclusions about measurement and use of health outcomes measures from the consideration of the full range of the four condition/treatment areas.

3.1 Elective Surgery

Recommended Measures

As part of its work on health care outputs and productivity the Department of Health commissioned a study from the London School of Hygiene and Tropical Medicine (LSHTM) and the Royal College of Surgeons of England (RCS) of the use of PROMs within elective surgery for cataracts, knee replacement, hip replacement, hernia repair and varicose vein repair. The OHE Commission has built on the findings of the LSHTM/RCS research. One of the main authors of the study, Professor Nick Black, is a member of the OHE Commission.

We have also discussed with BUPA its experience, stretching back nine years, of using PROMs for elective surgery in its hospitals, and with CHKS the pilot programme it initiated in 2006 with four NHS Trusts to collect PROMs data. Further information on BUPA’s work is in Box 3.1 and on the CHKS pilot programme in Box 3.2.

Box 3.1: BUPA's experience with PROMs

BUPA's approach to PROMs has focused on clinical governance as the key to improving both efficiency and safety. PROMs are therefore seen as one of a large number of elements designed to engage clinicians in continuous quality improvement. BUPA stresses that its focus is on positive learning rather than punitive performance management, and PROM tools have been chosen to maximise the relevance of the data to clinical decision-making.

BUPA started with the generic SF-36 measure of health and has progressively moved towards the use of condition-specific tools where a consensus has emerged that the latter are more sensitive. The risk of wasting resources and goodwill on scrutiny of false-positives dictates careful attention to statistical significance, so it has picked a few procedures of high volume and materiality, looking to these as sentinels for their specialties. This focus helps to keep the cost-efficiency high.

BUPA's work with Outcome Technologies, a BUPA subsidiary company who provide the collection and analysis services, has resulted in a simple process that stakeholders consider works well. Outcome Technologies continues to provide PROMs services to Spire Healthcare (recently BUPA Hospitals) and other hospitals.

Box 3.2: Patient-reported outcomes pilot programme from CHKS (2006 onwards)

CHKS, a commercial provider of analytical benchmarking for the NHS, has launched this programme to help NHS Trusts meet the patient choice agenda, and enable hospitals to see how successful a treatment is from a patient viewpoint. CHKS is attempting to develop a database covering most inpatients and recording patient characteristics, patient activity and patient-related outcome measures. CHKS are currently working with seven hospitals in four Trusts to collect patient reported outcome data using the SF12 and EQ-5D tools. All eligible electively admitted patients receive a questionnaire prior to admission and a follow up three months post discharge. At three hospitals information is also being collected for patients admitted as an emergency, three months post discharge.

CHKS is already collecting patient data from these hospitals, which allow it to link the outcome data back to the patient record so that all patients have an outcome score and this can be aggregated and reported by condition, procedure, HRG, specialty and consultant. The work started in April 2006 and is expected to run for at least three years. The experience of CHKS with implementing patient-reported outcome measures in health care institutions is important for informing the development of pragmatic measurement of outcomes within the NHS.

In the research commissioned from LSHTM/RCS by the Department of Health, the first step was to review the PROMs used in five elective surgical procedures and recommend a disease-specific measure and a generic measure for each procedure based on the measures' psychometric properties as well as their practicalities. The measures would then be piloted to explore the feasibility and practicalities of data collection, analysis and presentation. As a result, the following disease-specific measures were recommended for collection in four of the five surgical procedures:

- Hernia repair: none currently available
- Hip surgery: Oxford Hip Score
- Knee surgery: Oxford Knee Score
- Varicose vein surgery: Aberdeen Varicose Vein Questionnaire

Pilot testing of the Visual Function 14 (VF14) instrument revealed some serious concerns about its validity and the need for its modification or the development of a new instrument before routine assessment of the outcome of cataract surgery is implemented in the NHS.

Alongside the disease-specific measures, it was recommended that the EQ-5D index more (see Box 3.3) be collected as the generic measure. The EQ-5D was preferred to another generic measure, the SF-6D, as patients need more time to complete the latter, there is higher cost associated due to the need to pay a fee to the copyright owners and it has not yet been validated for this purpose in the UK. In cataract surgery no generic measure is recommended as neither of these generic measures included a domain on visual issues and both are therefore unresponsive to treatment outcomes for this type of surgery.

Practicalities of Using PROMs for Elective Surgery Procedures

The practicalities and the feasibility of PROMs collection are common to all four elective procedures and are likely to have relevance for other therapeutic areas. For an assessment of the value of a procedure, PROMs have to be collected pre-operatively and post-operatively. When a patient has an elective procedure it is usual for them to undergo a "pre-op" assessment. Therefore there are two options of when to administer the pre-op questionnaire: either at the assessment or on admission immediately prior to undergoing the surgery itself. There are advantages and disadvantages to both.

The researchers recommend 'on admission' for several reasons. At the pre-op assessment stage, it is not always known when the surgery is going to take place or even whether the patient requires the surgery until all the pre-op test results have been gathered. The surgery may be performed at a different centre and this can skew data on performance. In addition, patients may feel that their response may influence the decision on surgery and this bias may invalidate their response. Although recruitment on admission might be associated with higher levels of stress, and there is time pressure on staff at this particular stage, high levels of recruitment were achieved at the admission stage in the pilot programme.

The post-operative questionnaire is best administered through a central facility from where the questionnaire is despatched and where the responses are collated. This centre would be administratively separate from those performing the procedures, thereby reducing bias, i.e. patients may be less likely to feel that a negative response would be fed back directly to the surgeon and have any impact on their subsequent care.

Box 3.3: The EQ-5D

EQ-5D is a standardised instrument for use as a measure of health outcome. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status. It is designed for self-completion by respondents and is intended to be suited for use in postal surveys, in clinics and face-to-face interviews. It is cognitively simple and should take only a few minutes to complete. Instructions to respondents are included in the questionnaire.

The EQ-5D descriptive system comprises 5 dimensions of health:

- mobility
- self-care
- usual activities
- pain/discomfort
- anxiety/depression

Each dimension comprises three levels:

- no problems
- some/moderate problems
- extreme problems

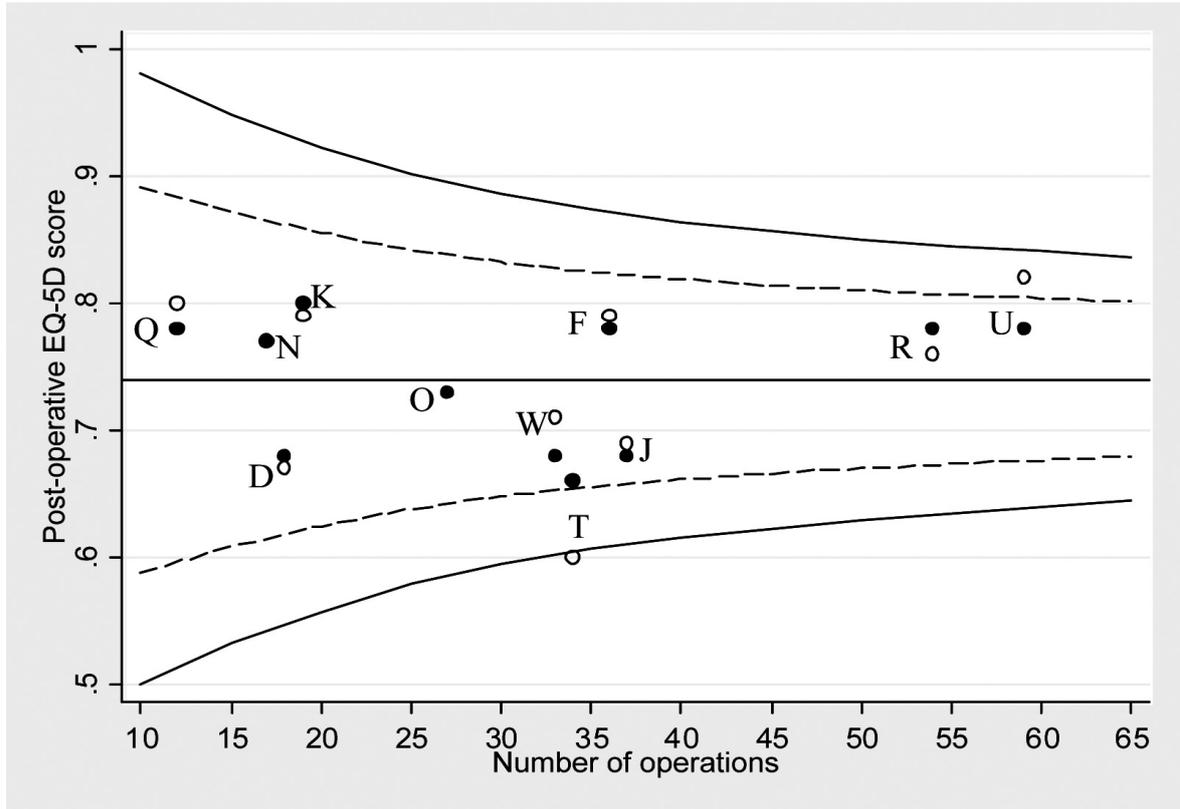
A unique EQ-5D health state is defined by combining 1 level from each of the 5 dimensions.

Source: <http://www.euroqol.org/>

The researchers consider that nursing staff would be better than clerical staff to recruit participants as they have greater experience interacting with patients. However, there would need to be training. Additionally enough time needs to be set aside for the patient to complete the questionnaire.

To enable PROMs to be used as performance measures at the local level, findings need to be presented in an easy-to-interpret way. Funnel plots showing outlying levels of outcome, above or below statistical confidence limits around the average outcome achieved nationally, are recommended for all potential audiences. Figure 3.1 shows an example for hip replacement operations: in this case all providers lie within the confidence limits which narrow as a unit undertakes more procedures per time period. Presentation of data in an acceptable way to all stakeholders is paramount to inform decision-making and is essential to the success of the use of outcome measures in the NHS, both for commissioning decisions and to give an accurate assessment of NHS productivity. At present, comparisons of PROMs can only be made at centre level rather than surgeon level for elective surgery, as the surgical procedure relates to the whole team rather than a single surgeon, and to enable sufficiently large samples to be collected in a timely fashion.

Figure 3.1: Funnel plot of post-operative EQ-5D scores in hip replacement surgery



○ unadjusted
● casemix adjusted

Source: Browne et al., 2007

Disease Specific and Generic Measures

In all the procedures the disease specific measure is found to be more responsive than the generic measure. The correlations between the disease specific measures and a generic measure such as EQ-5D have been shown to be either weak or moderate (see Section 3.5). Therefore the recommendation is in each case to collect both the disease specific measure (except for hernia repair) and the EQ-5D generic measure. There are confounders that need to be adjusted for, so variables such as age, sex, general health status and co-morbidity have to be collected alongside the PROM measure to be able to compare across centres.

The LSHTM/RCS research looked at three different ways in which the performance of the different centres could be assessed, i.e. through the change in the disease-specific PROM, the change in EQ-5D and the proportion of patients reporting complications. Patient-reported complications were included as both patients and clinicians regarded this issue as particularly important. While the validity of the absolute rates is uncertain, patient-reported complications can be used for comparing providers. The EQ-5D was found to be less responsive in the more minor procedures to the effects of treatment but it was felt that generic measures still need to be collected as the only way to compare across different surgical procedures.

The remit of the LSHTM research did not cover methods of incorporating any of these measures into a method of measuring productivity or an analysis of the time series of any outcomes data currently available.

PROMs in elective surgery may be used in other ways than those explored by the questions posed by the OHE Commission. In surgery there may be a threshold of quality of life. If a patient is above this threshold pre-operatively then their capacity to benefit may be so low that surgery is not worthwhile. However, PROMs have not been validated for this purpose yet. If such a protocol was instituted, gaming by patients may be a problem in response to the knowledge that PROMs are used to investigate the appropriateness of a procedure. Although PROMs may not be easy to use to evaluate the appropriateness of an intervention, the data are still useful for exploring trends. PROMs data may be particularly useful where patients could be helped to choose between different options and pathways if they had better information on the outcomes likely to result from the alternatives (The Dartmouth Atlas of Healthcare, 2007).

Time Series of Data

BUPA has been collecting outcomes data from patients at its hospitals for nine years – using the SF-36 in most cases and the VF-14 measure for eye patients. The NHS generally lacks similar time series of data on either a generic or on a condition-specific basis. However, data are available for some specific operations such as on cataracts – these show that the pre-op visual function of patients is significantly less severe now than it was in the 1990s (Black, personal communication). Surgical mortality rates are available for a number of years – but their usefulness is limited as a measure of patient outcomes given the relative rarity of surgery-related death. Data on mortality say nothing about the degree of success or otherwise of treatment for the large majority of elective surgery patients.

Changes in Mixes of Treatment and Costs

We are not aware of any studies of how the mix of types of treatment has changed for patients with conditions currently treated surgically and how this may have impacted on measured NHS productivity. To the extent that surgical procedures are over time delivered more efficiently, for example by reducing hospital lengths of stay, then traditional activity-based efficiency measures will record that. Where productivity measurement problems arise, however, is when the setting of treatment changes significantly – e.g. from an inpatient to an outpatient setting – or when improved approaches lead to looser referral thresholds so that more patients are being treated, including frailer patients, or patients with less potential for improved quality of life (e.g. because their condition is relatively minor), than previously.

In the first case, a change in the treatment setting, activity-based indices suffer from the weakness that it is difficult to assign weights: activity may be cost-weighted in which case movement of treatment location from, say, an inpatient to an outpatient setting would be recorded as a *reduction* in output even though the outcomes may be unchanged and only the cost reduced. What is needed is a measure of the outcomes achieved for the patient which is independent of setting.

When improvements in care due to new health care technologies or better use of existing technologies over time (learning effects) mean that sicker patients can be treated than previously, this may mean greater costs per patient, e.g. because of the complications resulting from their comorbidities. But it may also mean greater (or smaller) outcome improvements per patient. In that case, assuming that 'a patient is a patient is a patient' does not lead to a good estimate of the productivity change being achieved. Using a measure of outcomes achieved for each patient, individually, overcomes that problem.

3.2 Mental Health

Mental health differs somewhat from the other therapeutic areas we have explored as the link between interventions and outcomes may be less straightforward. Maintaining health status – i.e. preventing further deterioration rather than improving on current health levels – can more often than elsewhere be seen as a positive outcome. Measuring the outcomes of mental health services is an important aspect of outcomes and productivity measurement; they account for over 10% of total NHS spending (£8.5 billion in 2005/06 in England alone (Department of Health, 2007a)).

There has been a massive policy drive towards the use of routine outcome measurement in mental health starting in 2002 with the establishment of the Outcomes Reference Group, which eventually led to the publication of the Fonagy Report (Fonagy et al, 2004). This report stated that it was essential to collect local outcomes data to inform service delivery, and implementation should take account of the existing IT systems. Morbidity and mortality measures should be first priority, but the HoNOS (Health of the Nation Outcome Scales – see below for more detail) should be used as an anchor for other outcome instruments. The collection of HoNOS data was recommended to be made a requirement as part of the Mental Health Minimum Data Set (MHMDS). It was later made mandatory by the Department of Health. Four pilot sites ran the collection of the HoNOS data and it was felt that additional data such as diagnosis and interventions were necessary to supplement outcome measures to inform clinical decisions.

After the Fonagy Report there was a lull in policy measures focussed on outcome measurement in mental health until recently when the Care Services Improvement Partnership (CSIP), the umbrella organisation of the National Institute for Mental Health in England (NIMHE), instituted a national outcomes project. A compendium of mental health outcomes has been developed which will outline the properties of different instruments and when they may best be used (www.csip.org.uk/resources). However, the CSIP will not mandate these measures as they may be subject to changes over time. HoNOS is still the only measure mandated, but, despite this coverage is not as good as envisaged. The compendium is a way of introducing new measures which NHS organisations may want to use to drive service improvement or innovation.

The five year review of the NHS mental health National Service Framework (NSF) was published in 2004 and listed key priorities. But it made no mention of outcomes measurement, other than indirectly with regards to the lack of sophisticated information systems.

Disease-Specific Measures within Mental Health

For a measure to be useful to incorporate into a productivity index it needs the following features:

- wide coverage
- applicable in a number of care settings
- routinely collected in clinical practice
- readily linked to activity data
- can potentially be converted to a generic outcome measure
- available as a time series

In mental health most rating scales are completed by clinicians rather than patients. This increases the administration and training costs. Gaming and inter-rater variance is another likely adverse effect from clinician led measures, but this has to be weighed against the problems with patient led measures particularly in those who are acutely unwell.

There are two disease-specific measures that contend for incorporation into a productivity index – one of which is clinician led and the other is patient-reported, which allows for a comparison of the issues.

'Health of the Nation Outcome Scales' (HoNOS) was developed by the Royal College of Psychiatrists' Clinical Research Unit. It comprises a questionnaire which takes 5 to 15 minutes to complete and is part of the MHMDS as stated above. Training is required for raters and the cost of this may therefore be an issue. It has been tested thoroughly and has been proved to be valid, reliable and sensitive to change. Although it is part of the MHMDS, the overall completion rate within mental health trusts averages only 9.5%, but time series data do exist due to the length of time it has been part of the MHMDS. HoNOS has been recommended by the Department of Health which apparently makes it an unpopular instrument with clinicians. Policy-makers and academics appear to hold self-rated instruments in higher regard than clinician rated measures like HoNOS which are by definition based on a medical model of mental health problems. There seems to be a difference between clinicians and patients in what they perceive to be the most important aspects of their care.

The research on mental health outcomes undertaken by Rowena Jacobs for the OHE Commission included a series of interviews with 28 policymakers, academics and NHS staff involved in outcome measurement. These yielded contrasting views on HoNOS, ranging from it being considered a very blunt instrument, to it being very good at measurement even in the acute stage of the disease, to being more of a strategic tool mainly beneficial at the aggregate level. Any outcomes measurement instrument in any therapeutic area is likely to be subject to differing views. HoNOS is the instrument with the widest coverage in mental health in the UK and is the only one with a formal mandate.

The other outcome measure which may be a candidate for incorporation into a productivity index is the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM), which is intended to cover patients in the community setting or involved in psychological therapies. It was developed in a different way from the HoNOS scale through a multidisciplinary team and is supported by a not-for-profit organisation. The CORE-OM questionnaire takes 5 minutes to complete and has been found to have reasonable reliability, validity, sensitivity to change and acceptability. This outcome measure is used in over 250 counselling services across the UK. All the data are compiled in the database CORE IMS, which covers 100,000 patients per annum. There is a time series of data available from 1999 onwards in the database but the data quality in the early years may need to be explored in greater detail.

As CORE-OM is a self-rating instrument, it can be easily completed by less severely ill patients and patients with personality disorders. Patients are more likely to engage willingly in the process when they can see the changes in their own health over time. There can however be language and cultural barriers that reduce completion rates for self-reported measures particularly in mental health. CORE-OM is widely used in psychological services.

No evidence was found to suggest which of these two instruments, HoNOS and CORE-OM, would be better to use in measuring patient outcomes from mental health services, or rather which instrument to use in which circumstances. But we note that CORE-OM has proved popular and workable in community based mental health services and that HoNOS has been recommended for use in more acute phases of illness, such as those requiring hospital based care, to the extent that its collection has been mandated by the Department of Health.

Generic Measures and the Possibility of Mapping

As discussed with the research on elective surgery, for resource allocation decisions to be made or for productivity aggregated across different therapeutic areas to be assessed, generic measures of patient outcomes, rather than disease-specific measures, need to be collected. Generic measures of health related quality of life, such as the EQ-5D and SF-36 among others, have been used in a large number of evaluations of interventions in the field of mental health. Taken overall, however, there is evidently unease among health professionals about the value of such generic outcome measures at the individual patient level. There have been concerns highlighted over whether generic quality of life measures would be sufficiently responsive to mental health interventions.

The SF-36 contains questions about mental health and social functioning, and one of the five EQ-5D dimensions concerns anxiety/depression. However, generic measures of health related quality of life give rather more weight to aspects of physical functioning. Consequently they have numerous questions that may be felt to be irrelevant by mental health patients, making the questionnaires unacceptable to respondents. Generic measures may be insensitive to detecting changes in health status at the level of the individual patient. Generic outcome measures may be better suited at least in mental health to identifying health status changes at a population level.

An alternative approach to directly collecting a generic measure of patient outcomes is to attempt to map the measured disease specific outcomes onto a generic measure such as the EQ-5D if there is an evidence base from same patients being evaluated simultaneously with both the generic and the disease-specific measure. This has not been carried out in many instances in the literature to date, and not at all for CORE-OM. The scant evidence available suggests that there may not be a strong overlap between disease-specific measures of mental health outcome and generic measures such as the EQ-5D.

Time Series of Data

As noted above there are longitudinal data for both HoNOS and CORE-OM. CORE-OM data is available from 1999 onwards but there is a concern over data quality, and there may need to be negotiations over access to the data. The HoNOS scores are available for a period of three years. The data quality of the first of these is acknowledged to be poor, and there is still a question over the coverage of the data even in the latter two years. On the plus side the data can easily be linked to data in the MHMDS and this may be useful in the construction of productivity indices in the future.

Changes in mixes of treatment and costs

A number of studies have been published in the US looking at price indices for treatments within the area of mental health which specifically look at the change in treatments over time and the effect of this change on health outcomes. Studies have been published in major depression, schizophrenia and bipolar disorder, with a focus on the direct medical treatment costs related to episodes of illnesses. The studies suggest that costs of treating patients with these illnesses have fallen as a result of technological change, i.e. even though the total costs of care are rising, the patient outcomes are improving including the proportion of patients successfully treated. There have been shifts in treatment mix in all these areas. For example depression treatment has changed from primarily tricyclic antidepressants to serotonin selective reuptake inhibitors (SSRIs). The relative effectiveness of these treatments has been a matter of controversy. However Berndt et al (2000) in their study find evidence of falling average costs per successfully treated patient.

Technological changes in the US have therefore led to greater number of episodes treated, with similar or better effectiveness, at reduced cost per episode. Thus the increasing total expenditure on mental health services is due to increasing volume. More patient benefit is being obtained per dollar spent. It is uncertain whether the studies in any of these areas could be applied to the UK as the costing data available from medical claims in the US are not applicable to the UK NHS, but the MHMDS may be able to track changes in the actual treatment pattern. Evidence on effectiveness could be taken from one of the outcomes databases or from a published review of RCT evidence.

3.3 Chronic Obstructive Pulmonary Disease (COPD)

Treating patients for respiratory system problems cost the NHS in England alone £3.5 billion in 2005/06 (4.3% of total spend; source Department of Health, 2007a). Chronic obstructive pulmonary disease (COPD) is a major part of this. The Global initiative for chronic Obstructive Lung Disease (GOLD) defined COPD as '...a preventable and treatable disease...characterised by airflow limitation that is not fully reversible. The airflow limitation is usually progressive and associated with an abnormal inflammatory response to the lung to noxious particles or gases'. It is a major cause of morbidity and mortality, and it is increasing in prevalence.

Approximately 0.8% of the population has the disease but the proportion increases with age reaching 10% in men over the age of 75. The severity of disease is usually based on airflow measurement through the use of spirometry. The spirometer mainly measures the forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) for the designation of severity. As the disease gets worse, it is increasingly likely to be managed in secondary care with an increase in the use of oxygen therapy. If a patient has an exacerbation, it increases the likelihood of a hospital admission. Hence preventing the progression of the disease and stopping admissions to the hospital for exacerbations is paramount to the success of treatment.

Change in Mix of Treatments

The aim of treatment in COPD is to reduce the frequency and severity of exacerbations, improve health status and exercise tolerance, and to control symptoms. The severity of the disease is measured using spirometry. Treatment is recommended depending on the severity of the disease. For mildly ill patients short-acting beta-2-agonists and short-acting anticholinergics are usually used. As patients move to moderate severity a long acting bronchodilator either a beta-2-agonist or an antimuscarinic (e.g. tiotropium) is prescribed. Severely ill patients have corticosteroids added in combination with long-acting bronchodilators if they suffer from repeated exacerbations. Tiotropium and combination long-acting beta-2-agonist/corticosteroid (e.g. Seretide and Symbicort) are new additions to the range of therapies for COPD. Experts state that quality of life has been improved through the use of tiotropium and there is clinical evidence that mortality of COPD can be reduced by the use of the combination products. There has been a rapid uptake of both these types of products. Overall the drug treatment costs of COPD have increased from an estimated average of £15 per month in 2002 to an estimated £103 in 2007.

Other recent changes in management have been noted. Smoking has long been known as an important risk factor for COPD, and there has been an increase in the uptake of smoking cessation therapy. Oxygen therapy which is used in more severe cases has advanced through the advent of portable oxygen, and in diagnostic terms the use of spirometry has increased. The development of a number of treatment guidelines including those by GOLD, the British Thoracic Society and the National Institute for Health and Clinical Excellence (NICE) have contributed to the changing pattern of care.

Measuring COPD Outcomes

When a clinician assesses the progression of the disease they want to slow the decline of FEV1, whereas the patient simply wants to breathe easier. Clinicians though are starting to acknowledge the importance of quality of life rather than clinical end-points. The latter are collected routinely and include mortality, COPD related hospitalisations and COPD acute exacerbations. If an exacerbation does not require hospitalisation then it is more difficult to assess, although the GP research database (GPRD) may be able to provide primary care data and in future patients could be asked to respond to questionnaires administered in primary care settings or at home.

There are also a number of COPD specific clinical measures that are used in practice. We have mentioned FEV1 and FVC, but there are also gas exchange tests, haemodynamics and exercise performance tests e.g. the six minute walk test. There is no evidence of correlation between FEV1 and quality of life. FEV1 may not be relevant as a measure of overall health benefit. The gas exchange tests take too long to administer and do not have a specific health benefit component, whereas exercise performance could be informative if used in conjunction with other measures.

Disease-Specific Measures

A number of disease-specific measures have been developed to assess quality-of-life specifically in COPD patients. The researchers identified a structured review which undertook an assessment of generic and disease-specific patient reported measures within COPD. The review assessed five of them:

- 1 Chronic Respiratory Disease Questionnaire (CRQ)
- 2 St George's Respiratory Questionnaire (SGRQ)
- 3 Breathing Problems Questionnaire (BPQ)
- 4 Functional Performance Inventory (FPI)
- 5 Seattle Obstructive Lung Disease Questionnaire (SOLQ)

The CRQ was first published in 1987 and consists of four components: fatigue, dyspnea, mastery and emotional function. Patients need to identify areas of life that are affected by the disease which can make comparisons difficult. Sixteen papers were identified where the CRQ was used, and methodologically it appears to have sound properties. The SGRQ was published in 1992, and is a self-administered questionnaire for patients with airway obstruction such as in asthma or COPD. It has been used in a total of 12 studies, and again has sound psychometric properties. The other disease specific instruments may not be as suitable as the SGRQ and CRQ. The BPQ is not methodologically well supported. The FPI may be robust from a psychometric viewpoint but has not been used widely in practice. No evidence has yet been compiled around the precision of the SOLQ measure.

Disease specific measures have been tested in routine settings, but mostly in trials. Most of the trials have used the SGRQ either on its own or alongside a generic measure.

Generic Measures and Mapping

In the review of quality of life instruments used in COPD, seven generic instruments were identified:

- 1 SF-20
- 2 SF-36
- 3 SF-12
- 4 EQ-5D
- 5 Sickness Impact Profile (SIP)
- 6 COOP Charts
- 7 Nottingham Health Profile (NHP)

The SF-20 was developed in the medical outcomes study in the 1980s but only one study has been carried out since using it and only validity was proven out of the methodological properties. SF-36 was developed based on the SF-20, using the most important concepts from the earlier medical outcomes study, and other concepts used in other well-known health questionnaire to provide a comprehensive health survey. It was designed so that it can be self-administered over the phone, and has been widely used in trials. All psychometric properties have been demonstrated, and it appears methodologically sound in COPD as well as a wide range of other diseases.

The SF-12 is a shorter version of the SF-36, but designed to be equally as valid, i.e. it has the same health components. Although it has been proven to be methodologically sound, there is no specific evidence within COPD.

EQ-5D has been mentioned in other therapeutic areas within this report. It consists of five domains: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The questionnaire only comprises two pages therefore is relatively easy to fill in. It was identified in 25 COPD studies and there is significant evidence on its sound psychometric and methodological properties.

The SIP has been used in practice since the 1980s and consists of 136 statements about health related dysfunction. It can be self-administered or undertaken by interview process. Although it has been used in a number of COPD studies, there has been no recorded evidence of precision. Similarly there is insufficient evidence to support the psychometric properties of the COOP and NHP.

SF-36 is the most widely used of the generic measures. There is therefore much evidence to support its robustness. Both SF-12 and EQ-5D also fulfil all assessed psychometric properties, and all three have algorithms that can be used to convert it to QALYs, which is a necessary property to compare productivity across different therapeutic areas. Generic measures can monitor for changes in all domains of health, but the disease-specific may be more sensitive to small changes in specific components of COPD such as breathlessness.

There have been two successful attempts to map the disease-specific SGRQ onto EQ-5D, but no attempts in trying to map the CRQ to a QALY based measure. Two of the generic measures discussed earlier, SF-12 and SF-36, do not have associated utility values as with the EQ-5D. Brazier and colleagues developed the SF-6D for these two instruments so that a single index measure could be generated. An analysis has been carried out comparing the SF-6D to the EQ-5D, and the mean results were almost identical.

A national COPD audit was first put together by the Royal College of Physicians and the British Thoracic Society in 1997 (a limited study), and was repeated in 2001 (pilot audit) and 2003 (full audit). In the 2003 audit, data on admission, organisation of care and resources were obtained. Another audit is due in 2008 and will potentially monitor the change of resources and process of care. If health related quality of life could be routinely collected in this setting, changes in outcome and productivity over time could also be assessed.

The authors of the research undertaken for the OHE Commission proposed that mapping could be done from hard endpoint data such as admissions and exacerbations to HRQoL and health states. This could be done using published literature or through a simple research study. This would enable productivity measurement without additional data collection. However this mapped measure may only be sensitive to changes in certain aspects of the disease and treatment and is less satisfactory than direct measurement of outcomes.

Based on the findings just discussed, we propose the initiation of collection of generic patient outcome measures in this therapeutic area alongside disease specific measures, at regular intervals (e.g. during annual check-ups). This would enable a comparison of productivity within the therapeutic area and with other therapeutic areas.

Time Series of Outcome Measures

There are a number of databases which may provide time series data for COPD. These include the General Practice Research Database (GPRD), QResearch, and Mediplus in primary care and the Scottish Morbidity Board and Hospital Episode Statistics in secondary care. There are also three general databases which may be useful namely Health Survey for England, the Scottish Health Survey and the National COPD Audit. However these only include physiological data which is routinely collected such as lung function, mortality and acute exacerbations, but not health outcomes.

Health outcomes have not been routinely collected in clinical practice and therefore have only been longitudinally collected in clinical trials, for a maximum of three years. A short time series could therefore be constructed, but it may be simpler to map the hard endpoints for which time series data is available onto QALY estimates to obtain an approximate view of past trends in outcomes.

3.4 Colorectal Cancer

Cancers are, with cardiovascular disease, one of the two main causes of death in the UK, accounting for over one in four of all deaths. Cancer accounts for 5.4% of NHS expenditure in England (source: Department of Health, 2007a) and is the focus of a great deal of attention from government, the media and the public generally.

The Calman-Hine Report "A Policy Framework for Cancer Services" (Expert Advisory Group on Cancer, 1995) was a forerunner of the NHS National Service Frameworks. National policy on cancer has subsequently been taken forward through the NHS Cancer Plan (Department of Health, 2000) and the Cancer Reform Strategy (Department of Health, 2007d). NICE has been particularly active in the field of cancer developing service guidance on different cancers (the Improving Outcomes Guidance reports) and technology appraisals of new treatments. Decisions on the cost effectiveness of new anti-cancer drugs have had a very high profile in the media.

Box 3.4: Cancer outcomes

England has had comprehensive registration of all new cases of cancer for several decades. This has made possible monitoring of trends in incidence, survival and mortality for individual cancers. It has also enabled international comparisons to be made. The EURO CARE programme has highlighted the apparently relatively poor survival rates for many types of cancer in the UK in the 1980s and 1990s in comparison with other Western European countries. This has been a major driver for improvements in service delivery.

However, the usefulness of cancer registry analyses has been hampered by the lack of universal information on important casemix factors, such as stage of disease and comorbidity, and the lack of outcome measures other than death.

The Cancer Reform Strategy, published in December 2007, highlights the importance of collecting and publishing information both on patient experience and on clinical outcomes. Such information will be of enormous value to patients, who wish to make informed choices and to commissioners. Feedback of high-quality comparative data on clinical outcomes to clinicians and providers has already proved to be a major driver for improved quality in the NHS breast screening programmes. We now need to extend this to all patients with cancer.

The Cancer Reform Strategy and the NHS Operating Framework for 2008/9 put a duty on provider organisations to collect defined datasets for cancer. This will enhance the quality of information available to cancer registries and to national clinical audits. A new National Cancer Intelligence Network is being established to bring together all the necessary expertise (clinical, epidemiological and technical) to collate, analyse, interpret and disseminate this information.

Over time this will facilitate publication of casemix-adjusted cancer outcome data for all provider organisations in England.

Alongside this, and building on successful surveys undertaken in 2000 and 2004, regular surveys are planned of the experience of cancer patients managed within the NHS. These will enable individual teams to know where attention is most needed to improve the quality of patients' experience of care.

As survival rates for cancer patients improve, the quality of survivorship becomes increasingly important. A new National Cancer Survivorship Initiative is being established by the Department of Health, working closely with Macmillan Cancer Support and other charities. The development and implementation of patient reported outcome measures (PROMs) is likely to be high on the agenda.

Mike Richards

The OHE Commission decided to focus its attention on colorectal cancer for several reasons:

- Colorectal cancers are one of the four commonest cancers (along with breast, lung and prostate cancers. Each year around 36,000 new cases are diagnosed in the UK (2004 data, source: Cancer Research UK¹);

¹ <http://info.cancerresearchuk.org/cancerstats/types/bowel/incidence/?a=5441> accessed 24th October 2007. UK figure compiled by Cancer Research UK from government statistics from each of the four constituent countries of the UK.

- Colorectal cancers are the second commonest cause of death from cancer (after lung cancer), with round 16,000 deaths each year in the UK (2005 data, source: Cancer Research UK¹);
- Colorectal cancer has an intermediate prognosis with around 50% of patients surviving 5 years (2000-2001 data, source: Cancer Research UK¹). This compares with around 80% of breast cancer patients and only 7% of lung cancer patients surviving five years;
- Outcomes for colorectal cancer have changed considerably over the past 30-40 years. Only a quarter of patients diagnosed around 1970 survived for five years;
- There are several new options to improve outcomes from colorectal cancer. These include the introduction of a national screening programme, improvements in surgical techniques, the use of combined modality treatments (e.g. surgery plus radiotherapy plus chemotherapy) and the availability of novel treatments targeted against molecular changes in colorectal cancer.

For many patients with cancer their first concern is whether they are likely to live or die from the disease. However, as survival rates improve across a range of cancers issues relating to quality of life assume ever greater importance. For cancer it is therefore important to be able to measure both survival rates and quality of life. The distinction between disease-specific and generic measures of quality of life is then once more relevant.

Cancer Registries

Cancer is unique amongst the 'tracer' disease areas in that strenuous attempts are made to record all newly diagnosed cases through cancer registries. Cancer registration has been in place in England for over 40 years, enabling time series of data to be analysed. Registration of cancer patients is linked with the Office of National Statistics (ONS), to facilitate exchange of data when patients die. Cancer registries compile information on:

- Incidence, i.e. the number of new cases of each cancer reported in each year (which can be age standardised);
- Survival, i.e. the proportion of patients surviving for different periods after diagnosis (e.g. one year, five years and 20 years). One year survival rates are frequently taken to be a proxy measure for how advanced the disease was at the time of diagnosis. Five year survival rates are most commonly reported;
- Mortality, i.e. the overall death rate (e.g. per 100,000 population) from cancer;
- Treatments delivered in the first six months after diagnosis.

Cancer registries aim to collect information on the extent of disease at diagnosis (staging) as this is a critical casemix factor. However, at present this is not available for all cases.

Cancer registries do not currently routinely collect information on comorbidity or on local recurrence rates. However, pilots of such collection are currently being proposed.

Recent developments in information technology have enabled data from cancer registries to be combined with Hospital Episode Statistics. This will facilitate reporting of measures such as 30 day mortality following surgical interventions.

¹ <http://info.cancerresearchuk.org/cancerstats/types/bowel/incidence/?a=5441> accessed 24th October 2007. UK figure compiled by Cancer Research UK from government statistics from each of the four constituent countries of the UK.

The availability of international data on incidence, survival and mortality for individual cancers provides the opportunity for comparing UK outcomes with those achieved in other countries. Indeed, the findings from the EUROCARE programme, a collaboration between registries in over 20 European countries, have been highly influential in relation to national policy in England (Department of Health, 2000; Department of Health, 2007d).

Cancer registries do not currently record any patient reported outcome measures or any other quality of life measures.

The National Bowel Cancer Audit Project

The National Bowel Cancer Audit Project currently collects information on around 40% of incident cases of colorectal cancer. This project, which has been clinically led and is now supported by the National Clinical Audit and Patient Outcome Programme, provides some longitudinal analysis since 2000/01 of clinical outcomes such as 30 day mortality, complication rates following surgery and stoma formation. Attempts are made to collect information on stage of disease and comorbidity. But again there is no information on patients' quality of life.

Patient Experience Surveys

In England, the Department of Health and more recently the Healthcare Commission have undertaken a series of national patient surveys to obtain patients' experiences of their health care. One of these surveys, published in 2002, focused on assessing the quality of care as seen by hospital patients in 1999/2000 being treated for one of six types of cancer including colorectal. The total survey sample for all types of cancer was over 65,000 patients, being treated at 170 NHS Trusts. A 70% response rate was achieved, indicating the feasibility of undertaking large-scale postal surveys of cancer patients' experience of care. But while the survey provided a detailed 'snapshot' assessment of patients' attitudes to their cancer care at a particular time, it did not capture any robust evidence on quality of life or patient reported outcomes.

Disease-Specific Measures within Colorectal Cancer

The research we commissioned from the York Health Economics Consortium revealed a large number of patient outcome measures – both generic and disease-specific measures – being used in studies of colorectal cancer. Most of the measures found had been used in only one, or occasionally in two, studies. Five measures were found to have been used more frequently. In all cases the measures had been used in research studies. None was in routine use in a health care system.

The most frequently used instruments to measure patient outcomes in colorectal cancer were questionnaires developed by the European Organisation for Research and Treatment of Cancer (EORTC). The EORTC quality of life questionnaire for cancer patients, EORTC QLQ-C30, was cited in 29 studies found in the literature search. This measure captures information on patients' physical and emotional functioning, pain and symptoms, and a global assessment of quality of life. An adaptation of this instrument specifically for colorectal cancer patients, the EORTC QLQ-CR38, was cited in 15 studies. It adds to the EORTC QLQ-C30 eight questions specifically addressing colorectal cancer.

The only other disease-specific patient outcome measure to be cited more than twice in the literature review was the Functional Assessment of Cancer Therapy questionnaire designed for use in colorectal cancer (FACT-C).

Published literature indicates that both the EORTC QLQ measures and FACT-C perform well in terms of validity and responsiveness. The EORTC QLQ measures are also reported to be reliable measures, i.e. they repeatedly give similar responses in similar circumstances. No published evidence was found about the reliability of the FACT-C measure.

The ability to use outcomes measures routinely in the NHS depends on the:

- Acceptability of the measure to patients (especially ease and speed of completion);
- Practicalities of administering the measure, including staff willingness;
- Effort required to collate and analyse data.

The EORTC QLQ measures were found in several studies to be feasible for patients to use. In the large majority of cases patients found the questionnaires acceptable and were able to complete them unassisted. The FACT-C instrument was also found to be acceptable and there was some evidence to suggest that it could be completed in less time than the EORTC QLQs.

Interviews undertaken by YHEC with stakeholders in cancer services identified some local initiatives to capture patient outcomes data, including for colorectal cancer patients. These projects appear to have found a high degree of acceptance by patients, including the use of computer interfaces to complete the questionnaires (to reduce the burden of collation). However, adherence rates tend to fall over time, implying a need to avoid overburdening cancer patients with too-frequent requests to complete quality of life questionnaires.

Thus the introduction of colorectal cancer specific outcome measurement on a routine basis in the NHS looks worth pursuing.

Generic Measures and the Possibility of Mapping

The literature review undertaken for the OHE Commission found that generic health outcomes measures had been used with apparent success in one-off studies concerning colorectal cancer. But as for the disease-specific outcome measures, no generic measure is in routine use. The SF-36 questionnaire was cited in 10 studies found in the literature review and the EQ-5D in six. A number of other generic measures of patient outcomes had been used in single studies, e.g. the Health Utility Index and the Nottingham Health Profile.

Both SF-36 and EQ-5D have been found to be valid and reliable in a number of disease areas, with no evidence of inapplicability to colorectal cancer. One study (Wilson et al., 2006), compared the EORTC QLQ-C30 and FACT-C measures with the EQ-5D and the SF-12 (an abbreviated instrument derived from the SF-36) and found all measures to perform similarly well, although the EQ-5D may be less responsive than the other measures.

YHEC concluded that the SF-36, EQ-5D, EORTC QLQ and FACT-C measures may all be considered sufficiently robust for use in routine NHS practice. All the instruments have been widely used in research settings with relatively few concerns reported over their acceptability.

The introduction of generic measures of patient outcomes – SF-36 or EQ-5D – alongside, or instead of, disease specific measures which might be more sensitive to the impact of colorectal cancer treatments and more of interest to clinicians, therefore seems worthy of consideration. Use of EQ-5D would permit direct comparison of outcomes with those from other treatment areas, such as elective surgery.

Time Series of Data

None of the patient reported outcome measures discussed above has a longitudinal dataset available concerning patients with colorectal cancer. The only measures which permit longitudinal analysis are the mortality, survival and recurrence data recorded by cancer registries.

Changes in Mixes of Treatment and Costs

The research we commissioned has revealed no studies on changes in the mix of colorectal cancer treatments provided in the NHS over time, or on the cost consequences of such changes.

3.5 Using Disease-Specific Measures to Estimate Generic Health Outcomes

The detailed research undertaken for the OHE Commission on patient outcomes in elective surgery, mental health, COPD and colorectal cancer has revealed disease specific measures in use for research purposes in each case, but no usable time series of data yet in existence for the NHS. The usefulness of generic measures of patient outcomes, and particularly those leading to estimates of QALYs, has also been considered. Time series of generic outcomes data, other than mortality/survival data, do not yet exist.

Data on mortality have been collected for many years and so offer the possibility of trend data on outcomes. But although the mortality rate within, say, a year of treatment is an important patient outcome it is not very useful to assess the impact of health care in most cases. For the tracer disease areas discussed above, death soon after treatment is most useful as an outcome measure for treatment of colorectal cancer, though not even there if palliative services are under consideration. Thus, for most areas of health care, mortality data, though necessary and important, need to be supplemented by data on other outcomes.

Other generic measures of outcomes, and particularly the EQ-5D and SF series of measures, have been shown to be practical for routine collection from elective surgery patients alongside other measures specific to the type of condition/procedure in question (e.g. the Oxford Hip Score for hip replacement surgery). A combination of disease specific measures – such as the EORTC QLQ-38 – and a generic measure, particularly the EQ-5D, looks well worth piloting for colorectal cancer. For COPD patients we have found similar evidence: that valid, robust, acceptable, disease-specific measures of outcomes exist and have been used in research (e.g. the St George's Respiratory Questionnaire), and that generic measures such as the EQ-5D and SF-6D might usefully be collected in tandem. There is little evidence to date on using generic outcomes measures for patients receiving mental health services, and even the most commonly used disease-specific outcomes measures have met problems of low response rates when attempted to be used routinely.

If it were to be argued that it would be too burdensome to implement collection of both disease-specific and generic measures of patient outcomes, then a second-best approach that has been suggested is to focus solely on disease-specific measures and then use them not only in their own right but also as the basis for estimating generic measures. To help us to assess the desirability or otherwise of this approach, we asked Professor John Brazier and colleagues to review the evidence on mapping from condition-specific to generic health outcome measures, e.g. QALYs. Their full report is available on the OHE Commission website (www.ohe.org/page/ohecommission).

They found a modest but growing literature addressing this issue of "cross walking", as it is sometimes known; most of it published since 2000. The EQ-5D is the most common generic measure onto which other measures have been mapped. The extent to which generic measures

correlated with condition-specific measures was variable and not particularly strong. In nearly all cases the correlation coefficient was less than 0.5, implying that more than half of the variance between individuals' generic outcome scores could not be explained by variance between their measured condition-specific outcome scores.

The poor prospects for estimating QALYs, or other generic outcomes, from disease-specific measures contributes to our view that the preferred way forward should be direct collection of generic outcomes measures. Overall it appears possible to attempt routine collection of disease specific and generic measures together in many disease areas, although the introduction of generic health outcome measures in mental health services should probably wait until they have been tested in more promising fields. Our recommendations for introducing routine collection of outcomes measures are presented in Section 7.

4. RESEARCH FINDINGS – WIDER BENEFITS

Box 4.1: Wider benefits for patients and citizens

As patients we expect to receive reliable advice and effective treatment from well-trained professionals. We also want the security of knowing that health services will be readily accessible when we need them; that our views and preferences will be listened to by health professionals; that we will be given the help we need to help ourselves; that we can access reliable information about our condition and treatment; that the environment in which we receive treatment is clean and safe; and that we won't have to worry about the financial consequences of being ill. Furthermore, we want health professionals to show that they understand what it feels like to experience illness and undergo treatment, to anticipate our needs for information and support, and to treat us in a kindly and dignified manner.

What we want as individual patients and how we articulate these needs can be distinguished from our collective aspirations as citizens or members of the public. As citizens we may be concerned about more abstract notions of what constitutes a 'good' service, for example: affordability – free at the point of care; efficiency and value for money; universality, equity and fairness; safety and quality; health protection and disease prevention.

Measurement of these wider benefits is crucial for patients, citizens, health care providers and policy-makers. Patients need reliable information to help them select the best or most appropriate provider, treatment or care package. Citizens want to be assured that service quality is as good as it can be and that resources are being used efficiently and effectively. Health care providers need this information to inform quality assurance and improvement initiatives, and policy-makers need it to demonstrate accountability.

Much work has already been done on ways of measuring these wider benefits and surveys of patients' experience are now regarded as an essential component of performance assessment. The next challenge is to integrate patient experience and health outcomes measurement into routine data collection systems and to find ways to make the information accessible to all the different stakeholders.

Angela Coulter

4.1 Background

Benefits from health care can be thought of as falling into four categories, as shown in the following matrix:

	Health Benefits	Wider Benefits
Patients	1. (e.g. patient health related quality of life)	3. (e.g. patients informed about interventions and treated with dignity)
Others	2. (e.g. impact on carers' physical and mental health)	4. (e.g. benefits to the national economy from reduced time off work by patients and carers)

Most of the focus in the discussion of health care outcomes and productivity has been on the shaded cell 1 of this matrix: health benefits for the people treated. But part of the value obtained for the resources spent on health care derives from the other three categories of benefits.

The health of people other than the recipients of health care interventions may benefit from those interventions. The classic 'externality' of this kind is the benefit to their health that the non-immunised population receives when the people they come into contact with are immunised against infectious diseases. But in addition to that the health of carers, family and friends may be improved when someone is treated, e.g. for a chronic disease, due to reduced worry and stress on them (cell 2 of the matrix).

Patients may benefit from health care not only in the form of improved length and/or quality of life but also in other ways that can be thought of collectively as the humanity of the care they receive (cell 3 of the matrix). There are many dimensions to this kind of benefit and we discuss them at greater length below. It is also important to note that the humanity of care has an impact on patients' health, thereby contributing to the outcomes categorised in cell 1 of the matrix.

Then there are also numerous wider societal benefits which might result from health care interventions (cell 4 of the matrix), principally:

- Cost savings outside the health care system;
- Improvements to the productivity of the economy:
 - increased availability of patient and/or carer for work;
 - greater productivity while at work;
 - long-term future effects – if children are properly treated/vaccinated then there will be increased future productivity from them as adults, arising from less school absence as a child, less work absence as an adult or both;
- Public confidence or reassurance that comes from knowing that effective care will be readily available should it be needed. This is sometimes referred to as the "solidarity effect".

The wider benefits obtained by patients from health services (cell 3) include aspects as disparate as: the length of time spent waiting for services; their convenience to the patient; patients' knowledge and 'health literacy'; and the nature of patients' experience of the care process. Coulter and Ellins (2006) give examples of the last two of these:

- Patients' knowledge, for example:
 - Knowledge of their condition and of its long-term complications;
 - Self-care knowledge;
 - Knowledge of treatment options and their likely outcomes;
 - Comprehension of the information they receive about all of these;
 - Ability to recall this information;
- Patients' experience, for example:
 - Patient satisfaction;
 - Doctor-patient communication;
 - Psychological well-being;
 - Self-efficacy;
 - Patient involvement and empowerment.

Benefits attributable to health care but not directly about health improvement itself may be important both to the measurement of productivity of the NHS and to resource allocation decisions between services and providers within the NHS. Before these aspects of the outcomes of health care

can be given the appropriate weight in decision-making processes, we need to know their relative importance as compared to health outcomes.

The OHE Commission therefore commissioned a review of the WHO concept of health service 'responsiveness' as one way of capturing wider benefits to patients (cell 3); a review of evidence on wider societal benefits such as indirect cost savings at work (cell 4); and a review of whether published discrete choice experiments can give clarity to the relative importance of health outcomes and other types of benefits from health care.

4.2 Measurement of Responsiveness

As part of our analysis of benefits provided by the health care system which sit outside health benefits, we commissioned a piece of work from Nigel Rice and Silvana Robone from the University of York. Health system responsiveness is a concept promoted by the WHO and introduced by them with the release of their World Health Report 2000 (WHO, 2000). Responsiveness is defined as the way in which individuals are treated, the environment in which they are treated and the individual's experience of contact with the health system. A responsive health system will contribute to improve health outcomes, through access to services, adherence to treatment recommendations and more efficient use of health promotion information, but will also contribute to the other aspects of overall well-being identified by Coulter and Ellins (2006) and discussed above.

The quality of the interaction is of utmost importance and the responsiveness of the health care system is an input into this interaction. Responsiveness attempts to reflect actual experiences of contact with the system rather than satisfaction with the system.

The WHO has developed an instrument to measure health system responsiveness which was based on a literature review and informed the domains that would be covered in the instrument. These domains follow the principles:

- Validated in related fields as important attributes that individuals seek in their interaction with the health system;
- Amenable to self-reporting;
- Comprehensive;
- Measurable in a way that is comparable within and across populations.

The domains within the responsiveness measure defined by the WHO were:

- Autonomy;
- Choice;
- Clarity of communication;
- Confidentiality of patient information;
- Dignity;
- Prompt attention;

- Quality of basic amenities;
- Access to family and community support.

These dimensions cover essentially the same ground as the patient knowledge and patient experience dimensions described by Coulter and Ellins, referred to earlier. A responsiveness instrument was developed based on the eight domains and fielded in 71 countries as part of the WHO World Health Survey in 2002-2003. Respondents were asked to rate their most recent experience when filling in the survey (see Table 4.1). The survey consisted of a household questionnaire in the first section and the second section was at the level of the individual, looking at health states and the responsiveness of the health system. To date the responsiveness data have not been officially analysed.

There are problems with the measurement of responsiveness. It is particularly difficult to measure when investigating public health interventions. Higher responsiveness scores may be achieved in those systems that deny access and focus their attention on a small group of patients as opposed to a system that allows equal access to all.

Table 4.1: Example Questions Used in the World Health Survey to Measure Responsiveness

Domain	Question
Autonomy	How would you rate your experience of being involved in making decisions about your health care or treatment?
Choice	How would you rate the freedom you had to choose the the health care providers that attended to you?
Communication	How would you rate your experience of how clearly health care providers explained things to you?
Confidentiality	How would you rate the way your personal information was kept confidential?
Dignity	How would you rate the way your privacy was respected during physical examinations and treatment?
Quality of basic amenities	How would you rate the cleanliness of the rooms inside the facility, including toilets?
Prompt attention	How would you rate the amount of time you waited before being attended to?
Access to family and community support	How would you rate the ease of having family and friends visit you?

The domains of responsiveness developed by the WHO are for use for comparison across countries and at a more local level these domains may be supplemented for example an assessment of continuity of care may be of value when assessing mental health services.

Responsiveness is a multidimensional attribute, which once again brings to bear the problem of aggregation and how to weight responses from each domain. Weights are derived from a household survey which designates the importance of each domain. Common sets of weights used for

comparative analysis may introduce bias and therefore from a methodological viewpoint different weights may more appropriately account for the differing cultural and political factors. However the WHO has evidence that variability on all the domains is low enough to justify using common weights.

Differences in cultural expectations can also impact significantly on reporting behaviour. Hence attempts at developing an instrument that controls for individual preferences have been made, but there may still be different interpretations of the domain questions across subgroups of populations. Vignettes have been used to control for systematic differences across populations. Any systematic differences in responses to vignettes can be put down purely to differences in reporting behaviour, and these differences can then be related to the different characteristics of the different groups of respondents.

Both health benefits and wider benefits are thought of as appropriate outcomes to evaluate the performance of health care systems. Responsiveness aims to capture important dimensions of patients' experience. An instrument has been developed but has not been used much in practice. Therefore it is not known how the instrument behaves and whether it can be used comparatively to assess different health care systems.

4.3 Indirect Costs and Benefits of Health Care

Other types of outcomes that may result from health care include carer benefits and benefits to the wider economy. If treatments are successful, employees may be able to continue to work. These individuals may contribute to the economy to a greater extent than would be expected in the absence of treatment and this may have a bearing on the productivity of the NHS. The NHS contribution to national income may be underestimated if wider economic benefits of health care are not accounted for. Dr Clive Pritchard and Albert Chibi of the Office of Health Economics were commissioned to undertake analysis of how wider economic, i.e. 'productivity', benefits of health care could be measured and hence whether it would be feasible to analyse them from an NHS perspective. Their report is available on the OHE Commission website at www.ohe.org/page/ohecommission.

Three approaches for measuring these wider economic benefits are discussed in the literature. The human capital approach (HCA) values the loss of productivity attributed to disease. HCA looks at the amount of time individuals are able to allocate to work and is generally measured using gross wage, hence the effect of mortality uses gross wage from age of death to assumed retirement age. Those who have unpaid work have lost time valued as equivalent to those who have a similar paid job. There are many critics of the HCA. It tends to look at potential production losses rather than those that actually happen. This is because in practice short term absence may be covered by others temporarily or made up for on the employee's return to work. In the longer term also, work can sometimes be taken over by someone else at the organisation.

The second method for taking these issues into account is the friction cost approach, where the assumed existence of involuntary unemployment limits the production losses due to absence from work, i.e. those who are long-term absent or die can be replaced by someone drawn from the unemployed pool. This method does not recognise opportunity costs after the friction period, and therefore can underestimate production losses.

The third method is the "US panel" approach, where the inclusion of productivity costs is assumed to be dealt with by the measure of health effect. Thus individuals are assumed to take loss of income into account when they value their health states. This method can be thought of as the HCA method with productivity effects measured on the health benefits (e.g. QALYs) side.

In the Pritchard and Chibi review of the literature, 50 studies were identified: 46 of these focussed on productivity losses, seven on carer and care-giver costs, five on loss-of-leisure costs, two on unpaid work and two on schooling. Most of the studies did not state which method was used to calculate productivity losses. Out of those that did report, the majority stated that the HCA method was used.

When valuing productivity costs and informal care there is no consensus over which method should be used. There are many methodological debates ongoing about each method in different therapeutic areas. For example the human capital method assigns a zero value to neonatal mortality despite the psychological impact on the parents, and there is debate over whether illnesses such as depression would result in long enough absence to make replacement employment cost a probable issue. What is not doubted, however, is that productivity losses need to be investigated in a detailed and more systematic way. In some diseases treatment may have a greater effect on carers and working time, which are not accounted for when evaluating a treatment. Inclusion gives a truer picture of the productivity and performance of the NHS.

Independent of the method used, there may be ways of gathering data to use in calculation of economic costs. Days gained may be simply monitored using a health status questionnaire with an added question concerning time off work.

Pritchard and Chibi categorised studies according to the ICD-10 classification system of therapeutic areas. The therapeutic areas studied most frequently are listed in Table 4.2. No other therapeutic area had more than two studies.

Table 4.2: Therapeutic areas studied

ICD-10 chapter heading	Number of studies
Mental and behavioural disorders	12
Diseases of the nervous system	7
Diseases of the musculoskeletal system and connective tissue	7
Disease of the circulatory system	5
Diseases of the respiratory system	5

The most commonly mentioned condition overall was depression, with eight of the studies in mental and behavioural disorders falling into this category. Among diseases of the nervous system, the most frequently investigated condition was multiple sclerosis (MS) which was the subject of three studies, as was low back pain in diseases of the musculoskeletal system. In circulatory system and respiratory system diseases, interventions for stroke and asthma were evaluated by two studies each.

Pritchard and Chibi (see Appendix 6) found a study in the area of depression that argued that treatment was only 11.3% of the costs and a substantial amount of the rest of the costs could be attributed to productivity losses. In MS, one study reported that sufferers could lose up to 40% of their life time earning, and over 50% of patients have to retire within ten years following diagnosis. MS attacks in the most economically active age groups of the population, as does ankylosing spondylitis. Back pain is another major culprit of work productivity losses. 60% of employees have time off work due to this problem. In rheumatoid arthritis indirect costs were found to be much higher than resource utilisation costs, including those for informal care. In stroke, informal care constituted about a third of the overall costs. In both stroke and rheumatoid arthritis, the majority of patients are beyond working ages and therefore productivity losses may not be relevant but there are costs in the loss of leisure time as well.

The first two sub-sections of this Section of our report listed outcomes other than direct health improvement that may be the result from health care interventions, supplemented by work looking at a different framework for the patient experience, i.e. responsiveness. This third sub-section has looked at productivity losses at work resulting from diseases, and which are reduced by effective treatment of those diseases. It may well be worth exploring these losses for particular diseases, otherwise the wider societal gains from treatment may be underestimated. Does there, therefore, need to be routine collection of working days and informal carer data to inform productivity and performance measures? A number of diseases have already been identified where this would potentially be important, including depression and MS – these may be areas where it is worth piloting data collection. But before embarking on such a step it would be helpful to evaluate the relative importance of these wider societal benefits compared to the measured health outcomes. This is discussed in the following paragraphs.

4.4 Review of Discrete Choice Experiments

As part of the OHE Commission's investigations, we commissioned a review by Emma McIntosh (University of Oxford) of published discrete choice experiments to investigate the relative importance of different types of outcomes consequential to health care. A summary of her report is at Appendix 7 and the full report is on the OHE Commission website at www.ohe.org/page/ohecommission.

Many discrete choice experiments in health care focus on different aspects of health outcomes, but there are some that (also) focus on the process of care that may not necessarily be fully accounted for when measuring health outcomes. The number of published discrete choice analyses in health care is growing rapidly, and many of them have been conducted in the UK. They reveal a large number of characteristics being evaluated in addition to health gain, but among such attributes of health care those being reported most frequently as of significant importance to patients are:

- The length of time spent waiting for diagnosis and treatment;
- Continuity of contact with the same health care professional over time;
- The type of health care professionals delivering the care – e.g. doctors, nurses, therapists;
- The information provided to the patient.

4.5 Conclusions

A complete picture of the performance and productivity of health care services requires consideration not only of direct health care benefits for patients but also of the wider outcomes concerning the humanity with which health services are provided (including the responsiveness of the delivery of health care and the dignity with which patients were treated), health effects on others, and economic impacts outside the NHS. Pragmatism dictates a focused approach, identifying and measuring those wider benefits that are likely to be most significant.

National patient surveys already provide material concerning patients' experience of health services, in particular the national patient survey programme for England that is organised by the Healthcare Commission, and the Access and Choice survey carried out by MORI for the Department of Health. Other surveys do, or could, provide material of relevance, such as the GP practice surveys for the Quality and Outcomes Framework (QOF) and new surveys being planned by the General Medical Council and the Academy of Medical Royal Colleges for use in appraisal and revalidation of doctors.

With so many patient surveys in place or planned there is a need for better co-ordination to prevent duplication of effort. The Department of Health initiated a project in the autumn of 2007 to review the requirements of the health and social care system for "customer experience information". On

the back of that, we recommend that the Department of Health should work with its counterparts in Northern Ireland, Scotland and Wales on a UK-wide stock-take of patient surveys in process and planned, and then use this as the basis for discussion with all the interested parties how to co-ordinate those surveys in future.

The productivity impacts of health care on the economy are likely to be greatest in a relatively limited number of disease areas. The literature review undertaken for the OHE Commission highlighted depression, back pain, MS, stroke and rheumatoid arthritis, for example. We therefore recommend that after a short period of research to identify the most practical way of collecting such data, including working days lost, measurement of productivity effects be commenced for at least those patients being treated for the conditions just listed. Dependent on the success or otherwise of that pilot, collection of data might in the longer term be spread to all disease areas.

National surveys of a sample of the population – such as the Health Survey for England, Northern Ireland Health and Social Wellbeing Survey, Scottish Health Survey, Welsh Health Survey; and the General Household Survey in Great Britain and the Continuous Household Survey in Northern Ireland – might be one way of collecting at a general level data relevant to the wider societal impacts of health care.

5. AGGREGATION ISSUES

5.1 The Importance of Aggregate NHS Performance

The focus of the Commission's work has been on improving NHS performance. Can information on the things that matter to patients and taxpayers be collected in such a way that clinicians, providers, commissioners, the Department of Health and other devolved administrations, and ONS can understand whether the NHS is achieving better outcomes for patients?

As we noted in section 1.2 above, at the aggregate level we want to measure how NHS *outputs* as measured by activity, improve population health outcomes. Also, as we noted in section 4, there are wider benefits to patients beyond the immediate clinical outcome. The Atkinson Review expressed the issue in the context of the NHS and the National Accounts as follows:

"It is important to understand that we are interested in aggregate health outcomes not in order to develop a general measure of wellbeing or indeed of the overall 'health of the nation'. Such a measure would be interesting but is not the purpose of measuring NHS performance, which is to understand whether the NHS is getting better or worse at delivering health gain to patients. This includes aspects of *delivery* as well as of the health gain achieved by the intervention."

We noted that in preparing the National Accounts ONS is required by national, EU and OECD convention to focus on measuring *outputs* rather than *outcomes*. In the context of the NHS, output is the amount of care received by a patient, hence the use of NHS activity (i.e. what the NHS *does*) as a proxy for NHS output.

There are thus several attributes of measuring NHS health care output in the National Accounts that provide a challenge. These were expressed in the note (summarised at Appendix 9) we commissioned from Professor David Parkin (City University) as follows:

- it is provided publicly rather than sold in a market. Where services are not sold there is no natural unit of output "bought" by somebody;
- it is a customised rather than standardised product, with a complex production process. Measuring volume requires a unit of output, and this can be difficult to define if the goods produced are largely non-standard;
- the products that make up health care have primary and secondary characteristics. The dominant primary characteristic is the reason why the good is desired – an improvement in the person's health over the counterfactual of no health care. Secondary characteristics are also important: they affect the enjoyment of the primary characteristics of the good. In the context of the NHS the secondary characteristics are mainly related to how health care is provided.

For private sector goods ONS uses a mix of information on sales value, production volumes and prices to value output for the National Accounts. Understanding what is happening to the quality of goods and services over time is still a challenge. If prices go up does this reflect rising input prices or falling efficiency, both of which are bad news for the consumer, or an increase in quality which benefits the consumer? Hence in one important area of output – computers – adjustments are made by ONS to the price index to take account of the (improving) quality of the product. We referred in section 1 to the US literature on adjusting for quality in health care when measuring aggregate productivity. Given health care in the US is a marketed product, the US debate on quality in health care has been about how to adjust the *price* of health care services to take account of quality change over time. In the UK the debate is about quality adjusting output. The solution is the same in each case: measuring outcomes for patients.

5.2 ONS Work on Health Care Productivity: Using Outcomes to Adjust for the Quality of Activity

The ONS UK Centre for the Measurement of Government Activity (UKCeMGA) has, since the publication of the Atkinson Review, sought to take forward its recommendations across all non-marketed public services. It has done this, as has the Department of Health (2005, 2007e), by using data that are available *in the absence* of the routine collection of data on outcomes for NHS patients.

ONS published articles on health service productivity in 2004, 2006 and early 2008 (ONS, 2004; 2006a; 2008). All reported attempts to measure change in NHS productivity. The 2004 paper reflected a collaborative effort by the Atkinson Review team, ONS, and the Department of Health to improve NHS output estimates. Supporting analysis was provided in three papers in *Economic Trends* (Pritchard, 2004, Hemingway, 2004, and Lee 2004.) The 2006 and 2008 articles showed the effect of including some quality adjustments to the output measure. The article published in January 2008 had the following estimates for changes in NHS outputs and productivity over time.

Table 5.1: ONS estimate of NHS productivity change, 2001-2005

	Annual average change 2001-2005
Output growth	3.9%
Quality adjustment	0.5%
Quality adjusted output growth	4.4%
Input growth	6.5%
Productivity growth	-2.0%

Source: ONS, 2008.

In doing so it drew on estimates developed by the Department of Health (2007e). The two major contributors to the quality adjustment of 0.5% per annum were as follows:

- updates of the York/NIESR quality adjustment for hospital treatment, based on changes in 30 day survival rates, health gain following treatment for a limited number of procedures, and the impact on health gain of changes in waiting times;
- outcomes from primary care. The analysis included an estimated improvement in primary care productivity from changes in the proportion of patients with adequately controlled blood pressure and/or adequately controlled cholesterol.

The January 2008 article also notes the potential for using ONS data on trends in avoidable mortality over time. The issue will be to understand how much of this improvement can be attributable to the NHS.

The work by UKCeMGA and by the Department of Health is important but it simply does not make sense to continue to put considerable effort into working around the absence of data about the changes in health status and the mortality of patients receiving NHS treatment, when that data could be collected.

5.3 The Potential Benefits of Collecting Health Outcome Data

The York/NIESR Report included as its first main recommendation improving the quality adjustment in the measurement of aggregate NHS outputs and activity:

“Routine collection of outcomes data for a range of NHS treatments. The programme should start with a few high cost volume elective and medical conditions that would permit sampling rather than complete coverage. This data would also be immensely useful for other purposes including monitoring of Trust performance and improved cost-effectiveness analysis of particular treatments.” (Dawson et al., 2005, page 202)

Section 6 of the Department of Health paper “Healthcare output and productivity: accounting for quality change” Department of Health (2005a) set out 12 potential uses for systematic data on patient outcomes, of which the final one is “Measuring the quality gain provided to patients by the NHS, as part of the assessment of overall output and productivity.” The last two sentences of the paper conclude that:

“Use of existing data will be explored further. But more should be done to measure outcomes of treatment in terms of patients’ health status, either ‘before and after’ hospital admission or regular measurement in primary care, for patients with long term conditions.”

The ONS response (ONS, 2007) to its consultation on the way forward in adjusting NHS output to reflect changes in quality ONS (2006b) concluded:

“Consultation expressed a desire for better basic methods of measuring healthcare activities which linked separate episodes of treatment for the same patient. Data sets to measure patient health outcomes after treatment, routinely, would be very valuable. (paragraph 6.11)

This in turn reflected one of the comments of the Atkinson Report:

“It would be very helpful to be able to base quality adjustments for NHS output on a data set which measures the health outcome achieved as a result of treatment..” (paragraph 8.55, page 117)

There is no doubt therefore of the degree of support for the use of routine health outcome measures in the quality adjustment of NHS activity for aggregate measurement of NHS performance by the Department of Health and ONS. The OHE Commission’s view is that improving the measurement of aggregate NHS performance would not on its own justify a move to comprehensive collection of PROMs and of long term mortality data. Outcomes measurement is justified by the improvements to patient care that it will enable through better commissioning and better performance management. However, an important *additional* use of patient outcome measures will be to improve the measurement of aggregate NHS performance. The importance of having as full an understanding as possible of the productivity and performance of the NHS cannot be understated. ONS and the Department of Health have already made real progress but systematic collection of outcomes data is needed.

5.4 Remaining Issues

From the point of view of achieving a quality adjusted measure of NHS aggregate output, a number of the issues being explored by ONS, Department of Health and others will remain to be resolved even after the routine collection of outcome data is introduced. These include the following:

- If generic outcome measures are not collected then the use of data from condition specific instruments applied to some or all conditions would require either the development of algorithms to translate this data into QALYs – the “cross walking” we found is of limited validity – or the use of other techniques to offer a possible means to overcome the problem of incommensurable benefits. Professor Parkin has suggested to us that operational research techniques such as data envelopment analysis may offer a way forward in this area but they would require much developmental research to assess their feasibility;
- there are challenges in weighting different output elements to produce an overall quality adjusted output index. Instruments such as discrete choice analysis can be used to establish the relative weights of health outcome and key process factors such as waiting time for treatment (to the extent this is independent of any impact on health status) and the dignity with which people are treated by the NHS. But even within the area of aggregating changes in activity there is a common view across Atkinson, ONS, Department of Health and York/NIESR that it is important to move away from a *cost*-weighted index to a *value*-weighted index that better reflects the relative marginal value of activity to the marginal patient. This is for two reasons. Firstly prices not costs are the basis on which National Accounts are prepared. Secondly it moves us away from the perversity of reporting an NHS success of achieving the same or better health outcomes with a lower cost intervention as a reduction in cost weighted activity and so a *reduction* in NHS productivity and performance rather than an increase. Use of weights from generic PROMs may offer a way forward here;
- As we discussed in section 4, it is not obvious, in practical terms, how to account for the external benefits of health services, such as benefits to relatives of patients or improvements in workplace productivity. Issues include the extent to which these benefits may be regarded as being internalised by patients in their health state valuations. There is also, however, an issue of the complementarity between public and private output, for example higher real earnings mean that days lost through sickness have a higher value. The Department of Health has argued that there is a case for including “an estimate of 2% a year growth in the marginal social value of health output.” (Department of Health, 2005a, paragraph 10.17, page 63). There is a debate as to whether such an adjustment is relevant to valuing NHS output for aggregate performance measurement and for National Accounts purposes. ONS did not include it in its January 2008 paper on NHS output estimates;
- There may be lags, sometimes lengthy, between increases in inputs and increases in activity or outputs. There may also be services, such as public health, where estimation of outcomes is difficult. As Professor Parkin in his report to the OHE Commission (summarised at Appendix 9) put it: “it is hard to capture health improvement data from a general or patient population that is not in contact with the health system” although “it is plausible that population modelling of the impact of these activities on population health over time could provide reasonable estimates of future health improvements”;
- The Atkinson Review raises the issue of the ‘option value’ attached to the existence of a public service. There is a value to UK citizens from knowing the NHS is there, free at the point of use, even if he or she does not actually make use of the service in a given period. A change in this option value is a change in the value of the NHS.

- There is general recognition of the need for a disease-based approach. As the Atkinson Review put it "Ideally we should look at the whole course of treatment for an illness rather than at its components." (paragraph 8.39, page 113). An OECD/ONS Workshop in 2006 concluded that "Efforts should be made to measure "complete treatments" and the health gain arising from treatment" (Smith and Street 2007). Linking costs, activity and outcomes by disease as a route to build a picture of aggregate health system performance has long been advocated by Triplett, former Chief Economist at the US Bureau of Economic Analysis and an Advisor to the Atkinson Review and to the York/NIESR study (see for example Triplett, 1999). For three years, each PCT in the English NHS has prepared data on expenditure on healthcare across 23 'programmes' of care, based on the International Classification of Diseases (ICD) Version 10 disease categories. As noted by Martin et al (2007), "these programme budgeting data seek to allocate exhaustively to disease categories all items of NHS expenditure, including expenditure on inpatient care, outpatient care, community care, primary care and pharmaceuticals ..[and]..offer immense opportunities for examining the link between healthcare expenditure and health outcomes across PCTs." Routine collection of outcomes data, combined with programme budgeting data will offer the NHS a powerful tool to assess overall performance and to understand key areas of change.

There is thus much work to be done in this area, but the OHE Commission is clear that routine collection of outcomes data will transform understanding at the aggregate level as to whether and by how much the NHS is getting better at improving the health and related outcomes of the UK population.

6. IMPLEMENTATION ISSUES

Box 6.1: Commissioning for outcomes

There is now a consensus building that NHS commissioners should be focussing much more on the outcomes of care. However, it is clear from the analysis in this report that this is a significant task and that there is a process of experimentation and learning to go through before it can become widespread. For some conditions and providers it may always be the case that the outcome is too distant, the level of potential change too small or measurement not practical. In these cases commissioners will still need to use process measures and other proxies for outcomes.

Even where there is a full suite of outcome measures, commissioners will still need to take an interest in aspects of the process that are important to patients but which might not be directly connected to the clinical outcome. Indeed, in terms of the patient's perception, a successful clinical outcome may be partially or completely marred by a poor experience of the process. A number of the measures of wider productivity illustrated in section 4.1 will require a range of process information to be collected. The development of valid and reliable measures represents a challenge and despite the interest in localism it would better if there were national agreement on data definitions and standards wherever possible.

There is also an issue about how far it is legitimate for commissioners to be requiring highly detailed information about processes from providers. There is a strong view that commissioner-provider contracts should avoid putting the commissioner in a position where it appears to be micromanaging providers or creating a large and costly burden of data collection.

Although the methods are not perfect and there will be gaps where measurement will always be difficult, this is a classic case where the best should not be the enemy of the good. Starting to measure and use the results will, based on our experience with other areas, lead to a much more rapid improvement in data quality and measures than could be achieved by simply continuing to talk about it.

Nigel Edwards

The work of the OHE Commission has shown that the collection and use of outcome measures is paramount to enable:

- assessment of the productivity of the NHS as a whole;
- rational allocation of resources between different areas of treatment (e.g. acute elective surgery versus treatment of COPD); and
- as part of performance management and quality improvement, and patient choice, comparison of the outcomes achieved by different providers of the same services;

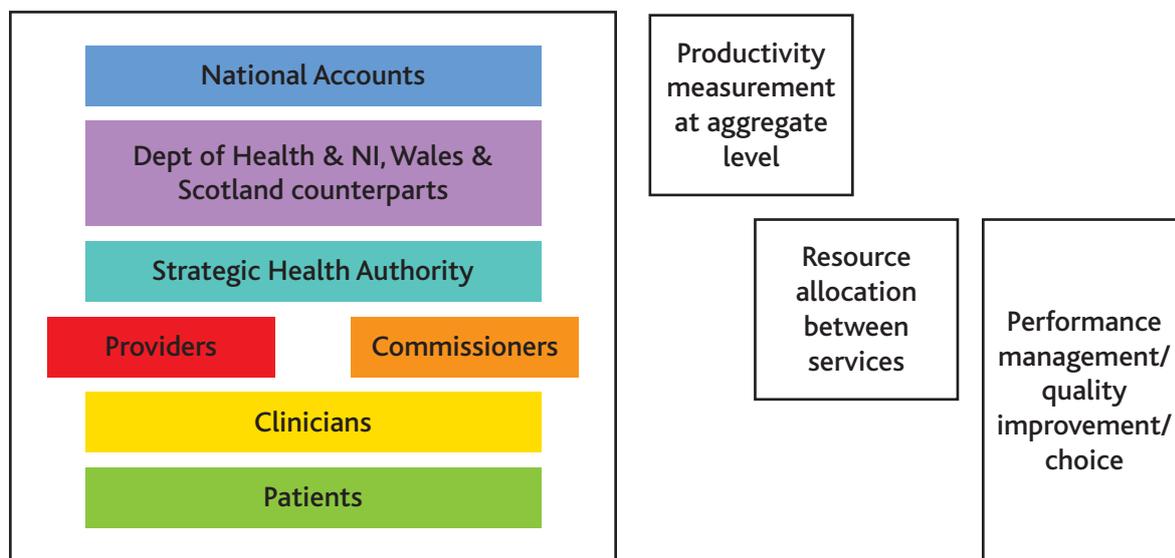
and thereby to achieve better care for patients.

Some useful measures of patient outcomes have been discussed in detail in earlier chapters. No matter how methodologically sound and useful these measures may be, unless the instruments are used and outcome measures collected, data will not be available to undertake appropriate analysis. It is therefore vital to involve the important stakeholders in the process from the beginning.

There are three main stakeholder groups that have to support and be involved in the process of outcome measurement for it to be successful (see Figure 6.1):

- clinicians
- patients
- managers (at both the provider and commissioner levels of the NHS)

Figure 6.1: Stakeholder Perspectives



6.1 Stakeholder Groups

Clinicians

If clinicians do not see the benefits of the measures then adequate data are unlikely to be collected in practice. A crucial barrier to the uptake of outcome measures would be if clinicians are unable to see the clinical benefit. The rationale for collecting these measures, including the benefit to clinicians' day-to-day work, needs to be clearly explained and linked to the principles of good clinical practice. It may be that clinician rated instruments may be more acceptable to clinicians than patient ratings, but as described in Section 3 there can be disadvantages to these types of instruments. Policy makers and patients may prefer patient-rated measures.

Where patient outcome measures are used, clinicians may prefer to use disease specific rather than generic measures but the two are not mutually exclusive: it is possible to collect both types of measure simultaneously. At the level of the individual clinician or clinical team, outcomes data will be valuable for performance management through quality improvement and clinical governance arrangements, revalidation, or to inform patient choice. Disease specific measures may be sufficient for that. But outcomes measures are also needed to inform resource allocation priorities, and to enable measurement of overall NHS productivity at the National Accounts level. It therefore makes sense to piggy-back the generic measures onto disease-specific measures to cover all purposes.

For patient outcome measures to be used in the first place and to be helpful in decision-making, clinicians must be involved in the process of implementation. This is most likely to be achieved if there are clinical champions to drive the use of outcome measurement, with the focus on enabling self-improvement: 'carrot' not 'stick'. This point was made strongly in Rowena Jacobs' findings on

measuring patient outcomes in mental health and in discussions at the October workshop we organised on commissioners' use of outcomes. A particular point made there was that even with financial incentive schemes, such as the 'Payment for Quality' adaptation of 'Payment by Results' being piloted in North West SHA's area, public and professional kudos for the clinical teams doing the work can be a stronger incentive than cash for the NHS Trust they work for. A central diktat from the Department of Health and the equivalent government bodies in Northern Ireland, Wales and Scotland can sometimes be counter-productive as the experience with HoNOS in mental health services indicates. Clinicians may resist a measuring process they feel is being imposed on them, unless its importance to good clinical practice is clear. Peer pressure can be a more effective means of increasing uptake: kudos for providers of health services who collect and publish their outcomes may be more of an incentive than any other.

Receiving regular and timely feedback is an important aspect for clinicians as well as the other stakeholders we discuss. Without this they will be unable to make decisions based on the data at the appropriate time, and so will see little point in collecting it. The data need to be presented in a way that individual clinicians, commissioners and providers can act on the results. The importance of appropriate IT infrastructure to enable this is clear.

To sustain the routine use of outcome measures there needs to be training in the use of these measures not only for the clinicians currently in practice but also as part of medical education for new doctors.

Patients

Informed patients provide an important lever to stimulate better outcomes from health care. Patient choice (using knowledge of outcomes to help choose between providers), patient driven quality improvement and outcomes-based commissioning appear likely to be strong drivers for the collection of outcomes data within the NHS. But if patient support is lacking, attempts to measure health care outcomes on a routine basis are unlikely to flourish.

The perspective of patients can be very different from that of clinicians. However there are some overlaps between the needs of the patient and those of the clinician. The questionnaire for assessing patient outcomes needs to be easy and quick to complete whether patient- or clinician-reported. In areas such as mental health patient characteristics could hinder the collection of data, and clinician support may be necessary.

Asking patients to complete questionnaires about their state of health should not be done lightly. There is a possibility that patients could suffer from 'questionnaire fatigue' but generic and disease-specific outcome measures can be collected within a single questionnaire to minimise the problem. To be sure about the patient outcomes being achieved by a particular provider for a particular service, and to enable reliable benchmarking between providers and reliable detection of trends over time, the ideal would be to try and collect outcomes measures from all patients treated. However, for services used with relatively large numbers of patients, it may be feasible to limit measurement of outcomes to those of random samples of patients, without excessive loss of reliability.

The use of outcome measures may be particularly useful to patients with chronic conditions such as COPD as they will be able to track the progression of their disease over time through their recorded health status. This should make patients more engaged, and may encourage a more active role, in their treatment.

Managers

There is management support for the use of outcomes as a means of measuring performance at provider and commissioner level. The availability of such data on a routine basis would aid performance management. This could be either providers analysing the performance of their own clinical teams, or commissioners managing the performance of the providers serving their patients. Managers are interested in using patient outcomes data to inform both the allocative decisions they make over which service/disease areas to invest more or less in, and quality improvement of the health care services provided. Outcomes measurement also helps managers and clinicians to see the impact of avoidable harm to patients, such as hospital acquired infections or other adverse events.

Making appropriate use of outcomes data is not an entirely straightforward task. For example, much of any observed variation in PROMs data over time or between providers at one point in time is likely not to be statistically significant, the result of 'noise'. Hence managers at providers and commissioners will need to tread carefully lest they waste time and goodwill on following false trails. Appropriate training in use – and how to avoid misuse – of patient outcomes data is likely to be needed.

However, an even more fundamental issue is that senior management teams have a number of priorities to pursue and outcome measurement simply must be near the top of this list to be taken seriously. It is important for collection not to be driven for the wrong reasons i.e. the need for clinical improvement must be the focus. If collecting, collating and analysing data on patient outcomes are seen only as 'tick box' exercises to prove that the correct governance arrangements are in place, then it is unlikely that measurement will be undertaken appropriately. Even if outcome measurement is centrally mandated, as has been the case with one of the mental health outcome measures, in practice implementation may be poor if its importance is not clear to all and the correct incentives are not in place.

There is an incremental cost involved in collecting patient outcomes data, an important consideration, but the magnitude is modest. BUPA have identified a cost of approximately £3 per patient providing information on the outcomes of elective surgery.

We recommend that measurement of patient outcomes and use of that information to drive better commissioning of services be made an explicit element of the government's vision of "World class commissioning" (Department of Health, 2007b). For outcomes measures to be taken forward, the outcomes data need to be useful, and/or incentivised, locally to both providers and commissioners. Only through use of these measures in practice will management see the benefits of outcomes measurement. Once again the data analysis and feedback needs to be timely.

We have briefly described the three main groups of stakeholders who will need to be the driving force behind outcomes measurement in the NHS if it is to be successful and their various perspectives. However, even if these stakeholders are fully engaged in the process, there needs to be good infrastructure and incentives to ensure that outcomes measurement can be achieved and useful within the NHS. As described in Section 3.1, two private sector organisations: CHKS and Outcome Technologies have already implemented systems to collect and analyse PROMs for, respectively, a small group of NHS hospitals and to a number of private hospitals. Such organisations, and no doubt other entrants to this market, can be called up for assistance, but even with their help much remains to be done.

6.2 Infrastructure

A number of the research reports we commissioned identified the need for good IT systems to support collection, collation and analysis of outcome measures. There is a debate over whether the information systems in place at the moment can handle the amount and rapid turn-around of data that is required. However while a unified NHS IT system is being put in place (Npfit), data could be collected and analysed across organisations. The IT systems need to be clinician friendly so that data are easy to input and analyse. Some resources will also be required for investment in training.

A patient friendly IT interface would be invaluable. There have been examples of touch-screen computer collection of health-related quality of life data in cancer patients. This proved to be less burdensome than the normal pen and paper approach. Such a system could be implemented across any therapeutic area. In some GP surgeries computers are currently used for electronic check-in, and could be advanced to include outcomes questionnaires for patients. In chronic diseases it would be particularly useful to administer a questionnaire at regular intervals at a check-up, e.g. annually. If the IT system is to be most helpful it should print out the results for the patient immediately and send them directly to the clinician to file in the patient's record.

Infrastructure does not only apply to IT systems. There is a shortage of resources currently in terms of people able to undertake the analysis of the collected data. Without these resources, data will not be fed back to the providers, commissioners and clinicians appropriately. If this feedback process is to work, there needs to be a framework in place over who should be responsible for the collection of data and how providers and commissioners are informed.

Commissioners do not yet routinely ask for health outcome data from providers when assessing their performance. It should be a matter of utmost importance to commissioners – primary care organisations and practice based commissioners – who need to drive the use of outcome measurement as they seek to judge in a more systematic way how their resources are used. However there is also an incentive for proactive providers to think about the information they collect to persuade commissioners to purchase their services. It seems reasonable for the responsibility to lie with providers as they are trying to inform commissioners and the performance managers of the benefits of their services. Commissioners need to have outcomes information from the providers who care for their patients, and hence commissioners also need an infrastructure of trained staff and IT to analyse and use that information.

6.3 Incentives

There are number of possible incentives that could be put in place to encourage the collection of outcome measures, using frameworks already in place in the NHS. We recommend that initially the emphasis should be on incentivising the collection, collation, analysis, feedback and use of patient outcome measures, rather than on rewarding providers according to the measured levels of outcomes achieved. In primary care, the Quality and Outcomes Framework (QOF) of the GP contract could be a good tool, i.e. GP practices could be paid for achieving high coverage in administering patient outcomes questionnaires that enable collection of outcomes data.

In secondary care, Payment by Results (PbR) may, in England at least, be another tool to incentivise uptake by offering premium prices to providers demonstrating the greatest assiduousness in collecting and using outcomes data. The 'Payment for Quality' pilot being undertaken by North West SHA is showing the way, although initially it is incentivising mainly process, rather than outcome, measures. But PbR may not be a solution in the short-term in areas, notably mental health, for which no tariff has yet been devised and applied, and outside England until activity based funding takes a hold there too. PbR-based incentives are likely to be modest financially and monetary incentives are

likely to have more effect on managers than clinicians. PbR provides strong incentives for the providers, but these tend not to be passed onto the clinicians hence they are managerial incentives. However, emerging experience of US 'pay for performance' schemes (upon which the North West SHA scheme is based) suggests that the professional kudos attached to being shown, publicly, to meet high standards can be a strong incentive for clinicians.

The financial aspects of both PbR and the QOF are not universally popular and the use of financial incentives needs to be piloted, monitored and evaluated rather than adopted wholesale without evaluation. Some feel that financial incentives may distort the care process and that kudos may be a preferable incentive, i.e. using outcome measures to show performance as a mark of a good quality provider. Outcome measurement is not currently a part of routine clinical practice and therefore there is a barrier to use in practice. Incentives need to be framed in a way to show that these measures are part of high professional standards. That could be aided by the collection of data on the outcomes of the care they provide being a consideration in the revalidation process for doctors.

Nationally promulgated guidelines for patient care may be another avenue for promoting the benefits of outcomes measurement, particularly if they are aimed at the commissioners of care as well as the providers. The Healthcare Commission already runs the The National Clinical Audit and Patient Outcomes Programme in the NHS in England, comprising clinical audits in a range of disease areas undertaken in collaboration with medical Royal Colleges and others, and responsibility for which passes to the newly formed Healthcare Quality Improvement Partnership in April 2008. The collection, collation, analysis, feedback and use of patient outcome measures need to be included in Healthcare Commission core standards (Healthcare Commission, 2006). The Audit Commission's proposed assessments of PCTs' use of resources could do likewise (Audit Commission, 2007). Such actions would encourage consistent and standardised measures being applied nationally which allow comparison and would reinforce the seriousness with which implementing outcome measurement is seen. At present, objective measures for clinical effectiveness are mentioned in the Healthcare Commission's developmental standards, but patient outcome measures are not mentioned specifically. Developmental standards are not mandatory at the moment but will be in due course.

We have heard how HoNOS measures in mental health although part of the MHMDS achieve coverage of only 9.5% of relevant cases, but there is no penalty for missing collection. Without having direct incentives, either a 'stick' or a 'carrot', it is unlikely that directives such as the MHMDS will work. Inclusion in the Healthcare Commission/Care Quality Commission annual health check of all NHS organisations would help to overcome this kind of problem by putting weight behind the implementation of outcome measurement.

A particular benefit of including outcome measurement in the Healthcare Commission standards would be its effect on application for Foundation Trust (FT) status. There have been examples where management have mandated a number of outcome measures to ensure that the Trust exudes a more business-like structure when applying for FT status. However, once FT status has been achieved these incentives become weaker and outcome measures become less of a priority once again.

6.4 Phasing and Pace of Implementation

The implementation process is constrained by shortages of skills and resources across the NHS but particularly amongst commissioners. But there needs to be a pragmatic implementation timeline. Outcome measures have to start being collected and used if their collection and use is to improve. Based on the research described in Sections 3 and 4 above, we conclude that while measurement of outcomes can and should be taken forward across a wide range of health services, the speed of implementation should vary.

The Department of Health has recently announced, with the publication in December 2007 of its "Standard NHS contract for acute hospital services" (for England) that from April 2009 providers must report patient reported outcome measures for: primary unilateral hip replacement, primary unilateral knee replacement, groin hernia repair and varicose vein procedures.

Similarly there is much that could be done to initiate the implementation of patient outcomes measurement in the other condition/treatment areas we have discussed earlier in this report. Specific recommendations including suggested timescales are set out in the next Section. Our general approach is to start as early as is practical in 2008/09 with measuring patient outcomes in a small number of diseases and procedures but depending on early experience with that to extend outcomes measurement, from 2010/11 say, to other disease and treatment areas as rapidly as resources permit over the following years. For example, if outcomes measurement proves practical in colorectal cancer it should be extended to other major cancer treatments. Based on experience in the initial service and treatment areas, patient outcomes measurement on a routine basis should subsequently be spread as rapidly as practicable to all areas of health care provision thereafter.

National patient surveys can be continued and adapted in the short term to permit collection of data on the wider impacts of health care and trends over time in economic impacts such as numbers of days off work through illness. It would be highly desirable also to instigate national surveys of carers to record the impacts on them of the health care received by the people they look after. Research into how this can best be implemented should be an early priority.

7. CONCLUSIONS AND RECOMMENDATIONS

The overall conclusion of the OHE Commission is that the collection and use of outcomes measures in the NHS is both practical and essential. We expect it to lead to improved outcomes, performance and productivity, thereby providing significant benefits to patients. In order to achieve the consequent benefits for patients sooner rather than later, we recommend an early start on that collection and use in a number of major disease areas and on a national scale throughout the UK.

In making recommendations, we are seeking a balance between speedy implementation rather than lengthy further consideration, but recognising that is not practical to apply outcomes measurement everywhere, at once, and allowing for refinements and more fundamental changes over time if proved desirable by emerging experience.

Although research has demonstrated the existence of statistically significant correlations between disease-specific measures of patients' health outcomes and generic measures of the same, that correlation still leaves a lot of unexplained variation. We conclude that there is no advantage going forwards in collecting only disease-specific outcomes measures, where generic health outcomes measures can be collected directly, as this involves little incremental cost.

Our recommendations, as discussed in preceding sections of the report, are set out below. In each case we have indicated the groups we consider should take responsibility for implementing them. Generally, the explicit support of the Government departments responsible for the NHS in the four countries of the UK, and the participation of other groups, including the health care professions, will be needed in addition to the groups named. Implementation is inevitably local to where patients are being treated, but a nationally consistent approach is needed.

N.B. We use the term "regulators" to include the Healthcare Commission, Monitor, the Audit Commission and the regulatory functions of SHAs in England, and the bodies undertaking those functions in Northern Ireland, Scotland and Wales. The term "providers" refers to all providers of NHS health care, including both primary and secondary care.

Recommendation

1. Measurement of patient outcomes and use of that information to drive better commissioning of services should be made an explicit element of the Government's vision of "world class commissioning".
2. Mortality and survival data should be collated and published for all condition/treatment areas at the same time as collection of other outcomes data is initiated. This requires data on casemix also to be collected, so that the crude mortality/survival data can be adjusted for the casemix treated. Confidence intervals should be estimated and reported to permit comparisons to be made over time and between providers. The same casemix data can be used to adjust other measures of patient outcomes too.
3. For elective surgery, condition-specific measures should be collected before and after surgery simultaneously with the generic EQ-5D for all NHS patients undergoing common procedures. Implementation should be at the earliest practical opportunity. In its "Standard NHS contract for acute hospital services" published in December 2007, the Department of Health has specified for the first time a requirement for providers to report (from April 2009) patient reported outcome measures for primary unilateral hip replacement, primary unilateral knee replacement, groin hernia repair and varicose vein

Action for:

Health departments

Commissioners
Providers
Regulators

Health departments
Commissioners
Providers
Regulators

procedures. We recommend extension to other common elective surgical procedures over the next three years unless emerging experience indicates otherwise.

- | | |
|--|--|
| <p>4. Measurement of mental health outcomes in secondary care should be re-invigorated with collection of HoNOS data being incentivised in addition to mental health providers being reminded of its mandatory nature as part of the Mental Health Minimum Data Set. We recommend that, in parallel, routine collection at annual intervals of the CORE-OM measure for all patients receiving community based mental health services be piloted in a range of NHS Trusts at the earliest practical opportunity in 2008/09. If the pilots are successful, collection of CORE-OM data should be rolled out to all providers of community mental health services to NHS patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers
Regulators</p> |
| <p>5. Further research is required to identify a generic measure of health related quality of life that is suitable for application to mental health patients as well as recipients of other kinds of health care. This may be based on an existing measure, such as the Health Utility Index (HUI3) or on adaptation of another measure such as the EQ-5D or one of the SF-series of instruments. The ultimate aim would be to develop a generic measure acceptable in all disease areas, including mental health. Research to that end should be funded by the National Institute for Health Research and its counterparts and commissioned in 2008/09 for completion within a year if possible. Subject to the results of that research, routine measurement of generic patient outcomes (e.g. at annual intervals) alongside the condition-specific measures could then be piloted in 2010/11 in a range of mental health service providers (both hospital and community based).</p> | <p>Health departments
National Institute for Health Research</p> |
| <p>6. The practicality and value of collecting and using measures of outcomes for COPD patients should be piloted by a range of commissioners and their secondary and primary care providers, commencing in 2008/09. The preferred disease specific measure is either the Chronic Respiratory Disease Questionnaire (CRDQ) or the St George's Respiratory Questionnaire (SGRQ). The selected disease specific measure should be collected simultaneously with a generic health related quality of life measure, for which the EQ-5D would be initially acceptable (pending the outcome of the research we propose in recommendation 5), so that outcomes could be compared with those for other conditions/treatments. If the pilots are successful collection should be rolled out to all commissioners and providers treating and/or managing COPD patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers</p> |
| <p>7. For colorectal cancer patients, mortality data (1 year and 5 years after diagnosis) should continue to be monitored and assessed as an important but partial measure of health service productivity in this disease area given its relatively high rate of mortality. But at the same time we recommend the instigation in 2008/09 of pilot programmes of routine use of the EORTC QLQ-CR38 instrument to measure outcomes in colorectal cancer patients simultaneously with the EQ-5D or an SF- series generic measure. Subject to the results of these pilot projects, collection of QLQ-CR38 and EQ-5D (or an adaptation of it) should be rolled out to all providers of colorectal cancer services to NHS patients from 1 April 2010.</p> | <p>Health departments
Commissioners
Providers</p> |

<p>8. In addition to patients' health outcomes, we recommend collection of data about the wider benefits from health services by addition of questions on patients' experience of care to the patient questionnaires used to measure health outcomes. Aspects covered should include: access, care co-ordination, autonomy, choice, communication, confidentiality, dignity, quality of amenities and support for carers. Whenever possible this should be done at the same time as the health outcome measures are introduced. Data on waiting times should continue to be used as a measure of one aspect of the humanity of care.</p>	<p>Health departments Commissioners Providers Regulators</p>
<p>9. Further research is needed into the relative values of the wider benefits referred to in recommendation 8 and by compassion with health outcomes. This should be commissioned during the course of 2008/9.</p>	<p>Health departments National Institute for Health Research</p>
<p>10. Information on wider benefits is already collected in national patient surveys. In the first instance we recommend that, building on its 'customer experience information' project, the Department of Health should work with its counterparts in Northern Ireland, Scotland and Wales on a UK-wide stock-take of patient surveys in process and planned, and then use this as the basis for discussion with all the interested parties how to co-ordinate those surveys in future. The stock-take and discussion should take place in 2008/9.</p>	<p>Health departments</p>
<p>11. After a short period of research to identify the most practical way of collecting such data, including working days lost, measurement of productivity effects should be commenced for patients being treated for mental health problems, diseases of the nervous system and diseases of the musculo-skeletal system. Pilot data collection should commence in the year 2009/10. Dependent on the success or otherwise of data collection in these areas, collection of data might in the longer term be spread to all disease areas.</p>	<p>Health departments National Institute for Health Research Commissioners Providers Regulators</p>
<p>12. Modification of existing national surveys of a sample of the general population – such as the Health Survey for England and its Northern Irish, Scottish and Welsh counterparts; the General Household Survey in Great Britain and the Continuous Household Survey in Northern Ireland – should be investigated as one way of collecting at a general level data relevant to the wider societal impacts of health care interventions, including impacts on carers.</p>	<p>Health departments</p>
<p>13. Health care professional bodies, including the Royal Colleges, should explicitly and actively support the collection and use by their members of patient outcome measures, including the provision of training in their collection and use for both existing and new clinicians.</p>	<p>Health care professional bodies Medical schools</p>
<p>14. Training in understanding and using outcome measures should be provided to NHS commissioner and provider managers.</p>	<p>Health departments</p>
<p>15. Investment should be made in IT and appropriately trained staff to enable user-friendly information collection, collation and analysis of outcome measures.</p>	<p>Health departments Commissioners Providers</p>

16. The NHS should, commencing as early as possible in 2008/09, pilot, monitor and evaluate explicit incentives to collect, collate, analyse, feedback and use patient outcome measures. Initially at least this would not be extended to rewarding providers for the measured levels of outcomes achieved, although that could be considered once experience with use of outcomes measures has accumulated. In primary care, the Quality and Outcomes Framework (QOF) of the GP contract could be a good tool, i.e. GP practices could be paid for achieving high coverage in administering patient outcomes questionnaires that enable collection of outcomes data. Other commissioners and providers might be incentivised financially through parallel arrangements. Publishing the outcomes collected, by organisation, will also be a vital part of the incentivising process. Good commissioners and providers will want to be seen to provide that information to the populations they serve.

17. There should be an expectation that within 5 years (i.e. by the end of 2012/13) routine measures of patient outcomes comprising the impact of an NHS intervention in terms of patient survival, quality of life, and experience of care, are collected for the majority of NHS activity. Within 10 years it should be the exception rather than the rule to find an area of NHS activity that is not assessed and actively managed according to the outcomes achieved.

18. The NHS should work with the ONS Centre for the Measurement of Government Activity to ensure that the health outcomes data used by the NHS can also be used by ONS in its periodic assessments of NHS productivity and, subsequently, in the National Accounts.

Health departments
Commissioners
Providers
Regulators

Health departments
Commissioners
Providers
Regulators

Health departments
ONS

REFERENCES

- Appleby A. and Devlin N. (2004). Measuring Success in the NHS: Using patient-assessed health outcomes to manage the performance of healthcare providers. A report commissioned by the Dr Foster Ethics Committee" London: Dr Foster. Available at: www.drfooster.co.uk/library/reports/measuringSuccess.pdf
- Atkinson A.B. (2005) "Atkinson Review of government output and productivity for the National Accounts: final report" Palgrave Macmillan: Basingstoke. Available at http://www.statistics.gov.uk/about/data/methodology/specific/PublicSector/Atkinson/downloads/Atkinson_Report_Full.pdf
- Audit Commission (2007) "Use of resources 2009. Comprehensive Area Assessment consultation" London: Audit Commission. Available at: http://www.audit-commission.gov.uk/Products/NATIONAL-REPORT/54B17295-BA67-454d-9255-88089F70BDA3/UofR2009_22Nov07REP.pdf
- Berndt E, Busch S, Frank R (2001) Treatment price indexes for acute phase major depression. In Cutler, D, Berndt E, (Eds) *Medical Output and Productivity*. University of Chicago Press. Chicago.
- Berndt E, Bir A, Busch S, Frank R, Normand S. (2002) The medical treatment of depression, 1991-1996: productive inefficiency, expected outcome variations, and price indexes. *Journal of Health Economics* 21(3): 373-396.
- Berndt E, Busch S, Frank R, Normand S (2005). Real Output In Mental Health Care During the 1990s. Working Paper 11557, National Bureau Of Economic Research. Cambridge, Massachusetts.
- Bridgewater B., Grayson A.D., Books N., Grotte G., Fabri B.M., Au J., Hooper T., Jones M., and Keogh B. (2007) "Has the publication of cardiac surgery outcome data been associated with changes in practice in North West England? An analysis of 25,730 patients undergoing CABG surgery under 30 surgeons in 8 years" *Heart*, July 2007;93:74-748.
- Browne J., Jamieson L., Lewsey J., van der Meulen J., Black N., Cairns J., Lamping D., Smith S., Copley L. and Horrocks J. (2007) "Patient reported outcome measures (PROMs) in elective surgery – Report to the Department of Health", December 2007; London School of Hygiene and Tropical Medicine. Available at: <http://www.lshtm.ac.uk/hsru/research/PROMs-Report-12-Dec-07.pdf>
- Coulter A. and Ellins J. (2006) "Patient-focused interventions – a review of the evidence" The Health Foundation: London. Available at http://www.health.org.uk/publications/research_reports/patientfocused.html
- Cutler D.M. et al (1998) Are Medical Prices Falling? *Quarterly Journal of Economics* 991-1024
- Cutler, D., McClellan, M., Newhouse, J. and Remler, D. (2001), "Pricing Heart Attack Treatments", in D. Cutler and E. Berndt (ed), *Medical Care Output and Productivity*, University of Chicago Press. Chicago.
- Cutler, D.M. and Berndt E, (2001) (eds). *Medical Output and Productivity*. University of Chicago Press. Chicago
- Cutler D.M. and McClellan M. (2001) "Is technological change in medicine worth it?" *Health Affairs* 20; (5): 11-29.
- Cutler D.M., Landrum M.B., and Stewart K.A. (2006) "Intensive medical care and cardiovascular disease disability reductions" National Bureau for Economic Research Working Paper No.12184, May 2006. NBER: Cambridge, Massachusetts.

Cutler, D.M. (2007) The lifetime costs and benefits of medical technology." *Journal of Health Economics*, 26:1081-1100.

Dartmouth Atlas of Healthcare (2007) "Preference-sensitive care" A Dartmouth Atlas Project Topic Brief. Dartmouth Atlas Project: New Hampshire. Available at http://www.dartmouthatlas.org/topics/preference_sensitive.pdf

Dawson D., Gravelle H., O'Mahony M., Street S., Weale M., Castelli A., Jacobs R., Kind P., Loveridge P., Martin S., Stevens P. and Stokes L. (2005) "Developing new approaches to measuring NHS outputs and productivity. Final report" University of York Centre for Health Economics Research Paper 6: York. Available at <http://www.york.ac.uk/inst/che/pdf/rp6.pdf>

Department of Health (2000) "NHS cancer plan: a plan for investment a plan for reform"
Department of Health: London.

Department of Health (2005a) "Healthcare output and productivity: accounting for quality change"
Department of Health: London. Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4124266

Department of Health (2005b) "Measurement of healthcare output and productivity. The routine use of patient reported health outcome measures (PROMs) in short-stay elective care environments" Technical Paper 1; December 2005; Department of Health: London. Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4124266

Department of Health (2007a) "Departmental Report 2007 – Department of Health" Cm 7093. The Stationery Office: London.

Department of Health (2007b) "World Class Commissioning – Vision" December 2007; Department of Health: London.

Department of Health (2007c) "The Operating Framework for the NHS in England 2008/09"
December 2007; Department of Health: London.

Department of Health (2007d) "Cancer reform strategy" Department of Health: London. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081006

Department of Health (2007e) "Further Developments in Measuring Quality Adjusted Healthcare Output. Department of Health: London. Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_079021

Expert Advisory Group on Cancer to the Chief Medical Officers of England and Wales (1995) "A policy framework for cancer services" (Calman-Hine Report) Department of Health: London.

Fonagy P., Matthews R., and Pilling S. (2004) "The Mental Health Outcomes Measurement Initiative: Report from the Chair of the Outcomes Reference Group" National Collaborating Centre for Mental Health: London.

Frank R.G., Berndt E.R., Busch A.B., and Lehman A.F. (2004) "Quality-Constant "Prices" for the Ongoing Treatment of Schizophrenia: an Exploratory Study" *Quarterly Review of Economics and Finance*. 44 (3):390-409

Healthcare Commission (2006) "The annual health check 2006/2007. Assessing and rating the NHS" The Healthcare Commission: London. Available at: http://www.healthcarecommission.org.uk/_db/_documents/The_annual_health_check_in_2006_2007_assessing_and_rating_the_NHS_200609225143.pdf

Hemingway J. (2004). "Sources and methods for Public Service Productivity: Health" *Economic Trends* 613: 82-90. Office for National Statistics: London.

Lee P. (2004) "Public Service Productivity: Health" *Economic Trends* 613: 38-59. Office for National Statistics: London.

Mai N. (2004). "Measuring health care output in the UK: a diagnosis based approach" *Economic Trends* 610: 64-73. Office for National Statistics: London.

Martin, S. Rice, N. Smith, P.C. (2007). The link between healthcare spending and health outcomes. Evidence from English programme budgeting data. QQUIP. The Health Foundation, London

ONS (2004) "Public Service Productivity: Health." Office for National Statistics: London. Available at <http://www.statistics.gov.uk>

ONS (2006a) "Public Service Productivity: Health." Office for National Statistics: London. Available at <http://www.statistics.gov.uk>

ONS (2006b) "Health Consultation. Measuring performance in our public services." Office for National Statistics: London. Available at <http://www.statistics.gov.uk>

ONS (2007) "Measuring Quality as Part of Public Service Output: Strategy following consultation." Office for National Statistics: London. Available at <http://www.statistics.gov.uk>

ONS (2008) "Public Service Productivity: Health Care." Office for National Statistics: London. Available at <http://www.statistics.gov.uk>

Pritchard A. (2004). "Measuring government health services output in the UK national accounts: the new methodology and further analysis" *Economic Trends* 613: 69-81. Office for National Statistics: London.

Smith S.C., Cano S., Lamping D.L., Staniszewska S., Browne J., Lewsey J., van der Meulen J., Cairns J., and Black N. (2005) "Patient-reported outcome measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence. Final report to the Department of Health" December 2005; Health Services Research Unit, London School of Hygiene and Tropical Medicine: London.

Smith P.C. and Street A. (2007). Report on Workshop on Measurement of Non-market Output in Education and Health, London, 3-5 October 2006. Centre for Health Economics, University of York.

Triplett J.E. (1999) "Accounting for Health Care. Integrating Price Index and Cost-Effectiveness Research". in Triplett J.E. (ed) (1999) *Measuring the Prices of Medical Treatments*. Brookings Institute Press, Washington DC.

Triplett J.E. (ed) (1999) *Measuring the Prices of Medical Treatments*. Brookings Institute Press, Washington DC.

Wilson T.R., Alexander D.J., and Kind P. (2006) "Measurement of health-related quality of life in the early follow-up of colon and rectal cancer" *Dis Colon Rectum*. 2006;49:1692-702

World Health Organization (2000) "The world health report 2000. Health systems: improving performance" WHO: Geneva.

Appendix 1 – Patient Reported Outcome Measures for Elective Surgery

A Report for the Department of Health

London School of Hygiene and Tropical Medicine and Royal
College of Surgeons Clinical Effectiveness Unit

EXECUTIVE SUMMARY

1. Feasibility

Where and when should patients be recruited?

According to both the preferences of patients and the recruitment rates achieved, there is little difference between pre-op assessment clinics and on admission. However, there are some disadvantages to using the pre-op assessment clinic so **we recommend recruitment on admission**. Around 80% recruitment can be achieved irrespective of surgical procedure or type of facility.

How can recruitment be maximised?

The main cause of non-recruitment was a failure to invite patients to participate. Overall, about 25% of potentially eligible patients were not invited to take part. Failure to invite in many centres is understandable given the short period of time in which staff had to adapt their routine practices to the study protocol. **We recommend a longer period in which to develop local arrangements, staff training to identify eligible patients and demonstration of the benefits of auditing patients' care.** We believe these steps will lead to high invitation rates. However, the active support of senior clinicians and managers is also needed, as are clear responsibilities and disciplined procedures.

Which staff are best suited to the task?

Nursing staff are better placed than clerical staff to administer questionnaires, possibly because of their experience in interacting with patients. With correct training, however, there seems no reason why clerical staff should not be equally proficient.

How can ineligibility be minimised?

The ineligibility rate (3.9%) was low for hernia, vein, hip and knee surgery. It was rare for a patient to be incapable of completing a written questionnaire in English (though few centres were in areas with a high proportion of ethnic minorities). However, up to 30% of cataract patients at some centres were deemed unable to participate due, not surprisingly, to vision problems. **We recommend investigation of the use of an interviewer to administer questionnaires to some cataract patients.**

How can patient agreement to participate be maximised?

There was a wide range of results across centres (33% to 100%) with a few very poorly performing centres skewing the overall mean rate (87%). **We recommend the questionnaires be shortened by excluding some questions that have been shown to be redundant, eliminating the research consent form which would not be required, and supplying the information sheet separately.** We believe that the agreement rate can be increased to an acceptable level in any future routine PROMs programme.

How can deceased patients be identified in a timely fashion?

To avoid causing bereaved relatives and friends any distress, we used the NHS Strategic Tracing Service to identify any post-operative deaths. This was successful in eight of the eleven deaths that occurred. **We recommend use of the NSTS** which enabled us to limit the sending of post-operative questionnaires to only three deceased patients out of a total sample of 2310 patients.

How can a high post-operative survey response be achieved?

In the light of the high post-operative response rates achieved (80-90%), **we recommend that post-operative questionnaires be dispatched and returned to a central facility, separate from the centre performing the surgery.**

What is the overall cost?

Based on necessary (see below) and achievable recruitment (80%) and response (80%) rates, the best estimate of the cost per patient was about £6.50. About 22% was met locally (staff time recruiting patients and using NSTS to trace post-operative deaths), making the central cost about £5 per patient successfully recruited. Almost 60% the central cost was for data entry, a factor that could be reduced using different technology.

2. Centre comparisons

Which PROMs should be used?

This project has confirmed the responsiveness of four disease-specific PROMs for assessing the impact of surgery. Concerns about the content validity and responsiveness of the VF-14 mean that a better instrument needs to be developed. We recommend the use of:

- **hernia repair: none (until a disease-specific instrument has been developed and tested)**
- **hip surgery: OHS**
- **knee surgery: OKS**
- **vein surgery: AVVQ**

Although it is possible to map changes in disease-specific scores onto a generic PROM (EQ-5D), the association for some procedures is weak (vein surgery) or only moderate (hernia repair). If the benefits of different procedures are to be compared, **we recommend the inclusion of a generic utility measure.** There is little difference in responsiveness between EQ-5D and SF-6D. Given the greater time involved in completing the SF-6D (36 items compared with 5 items), the higher incidence of missing data and the higher cost (given that use of the SF-36 scale normally requires payment of a fee to the copyright owners) **we recommend the use of the EQ-5D.** This does not apply to cataract surgery where both generic measures were unresponsive.

In addition, given their importance to patients and clinicians, **we recommend the inclusion of patient-reported complications.** It is important to recognise that the absolute incidence of complications reported by patients may include post-operative experiences that clinicians would define as 'normal' (eg wound discomfort) and also events that may not relate to the surgical operation (eg a urinary infection). In addition, the reliability and validity of complications measurement has not been tested as rigorously as other PROMs. The use of patient-reported complications should therefore be restricted to comparisons (such as between providers or procedures).

What are adequate recruitment and response rates?

Patients who are less healthy (as determined by the EQ-5D) are less likely to be recruited and less likely to respond to post-operative questionnaires. Post-operative non-responders also tend to be younger than responders. Considering the recruitment and response rates achieved at some of the participating centres, **we recommend 80% recruitment and 80% response rates should be sought to reduce any biases to insignificant levels.**

How should data be analysed to compare centres?

We recommend imputing missing items for all PROMs. The advantage of imputing is that it retains data and maintains sample sizes that would otherwise be reduced. The clustering of patients nested within centres did not affect the centre comparisons. **We recommend reconfirming the lack of clustering with a larger data set.**

Given the marked differences in case-mix (patients' pre-operative characteristics) between centres, any comparison of outcomes must take such differences into account if it is to be meaningful. We developed risk adjustment models for each PROM which explained a large proportion of the variance in post-operative scores observed. **We recommend the following confounders be included to risk-adjust post-operative PROM scores: pre-operative PROM score, age, sex, general health status, comorbidity (eight systemic conditions) and previous similar surgery.** Duration of symptoms is not required. The inclusion of the IMD (Index of Multiple Deprivation) was useful only for knee surgery. **We recommend investigation of the impact that ocular comorbidity might have on risk adjustment for cataract surgery.**

The models included all characteristics that it is feasible to collect from patients (rather than from clinicians), with the exception of height and weight. *We recommend improvement in these models by collecting much larger samples of centres and patients, and by the inclusion of some clinical factors (e.g. ocular morbidity for cataract surgery) either collected from clinicians or derived from existing clinical databases.*

Should 'minimally important differences' be employed to compare centres?

Both the anchor-based and distribution-based approach to generating MID values for PROMs are problematic: the former because a stable value, representative of the continuum of pre-op severity, cannot be generated; the latter because the value generated has no known relationship with patient experience. **We recommend that MID values be used with great caution when interpreting PROMs data and that statistical significance testing should remain the main approach to the comparison of centres.**

What power and statistical significance levels should be used?

Given the undesirable consequences of both false positives and false negatives in quality of care research, stringent criteria should be used with respect to both statistical significance and power when comparing centre performance on PROMs. **We recommend that statistical comparisons of centre performance should be based upon 95% power and a significance level of $p < 0.002$ (i.e. 99.8% confidence intervals).**

What is an adequate sample size?

We recommend a sample of about 150 patients from each centre to make meaningful comparisons. This would be sufficient to detect a difference of 1 Standard Error of the Measure (for both disease-specific and generic measures) with 95% power and statistical significance of $p < 0.002$ (equivalent to 99.8% confidence intervals) for cataract surgery, hernia repair and varicose vein surgery. For hip and knee surgery this sample would be adequate for the generic measure but for the disease-specific measures the power would be slightly less.

3. Summarising and presenting findings

What is the best way of presenting data?

We recommend the use of funnel plots that show both the unadjusted and adjusted data. This approach also allows more than one control limit to be shown, thus allowing the observer to impose their own level of certainty. This does not preclude also providing access to more detailed tabulated data, particularly for surgeons.

What is the best indicator for comparing centres?

We recommend using the mean post-operative PROM score adjusted for patients' pre-operative characteristics as the indicator of health gain. In addition, the proportion of patients reporting at least one complication (risk adjusted) should be used.

At what geographical level should centres be compared?

In view of the travel distance that patients are likely to find acceptable for common surgical operations, **we recommend that the performance of providers in the local 'region' be provided together with the national average.** Data for all providers in the country could also be made available, particularly for surgeons.

At what level of provision should comparisons be made?

We had sufficient data for making comparisons of centres, not individual surgeons. Despite patients and provider managers expressing an interest in comparisons of surgeons, current NHS policy on choice is limited to centres and the outcome of surgery relates to the performance of a whole team, not just the operating surgeon. Therefore, **we recommend that at present the unit of comparison is the centre.**

How frequently should performance information be updated?

As data collection systems become more sophisticated, it may be possible to provide continuously updated information. The feasibility of setting up such a system should not be underestimated. In the meantime, **we recommend 'batch processing' that updates every six months.** Allowing for three month post-operative outcomes, comparative information on providers for hernia, vein and cataract surgery would relate to performance 5 to 10 months previous. For hip and knee replacement, which currently have a six month follow-up, the information would be 8 to 13 months previous.

Appendix 2 – Investigating Patient Outcome Measures in Mental Health

Research Report for the OHE Commission on NHS Productivity

Rowena Jacobs, University of York

EXECUTIVE SUMMARY

1. This report examines the feasibility of incorporating patient outcomes in mental health into a productivity measure. It examines which outcome measures are most commonly used in mental health, the practical issues about collecting these outcome measures, whether they can be converted into a generic measure, whether there is a time series of data available, and whether the data exists to examine changes in the mix of treatments over time. The criteria that were assumed to be important for an outcome measure to be included in a productivity index, were that it should have wide coverage, should be routinely collected, could readily be linked to activity data, could potentially be converted to a generic outcome measure, and would be available as a time-series. The report focuses predominantly on mental health outcomes within the working age population. Literature searches on outcome measurement in mental health covered numerous databases and retrieved over 1500 records. Around 170 full papers were obtained.
2. Two measures emerged as the most likely contenders for inclusion in a productivity index in mental health, namely the Health of the Nation Outcome Scales (HoNOS) and Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM). HoNOS is used for patients with severe and enduring mental illness, while CORE-OM covers patients most often treated in the community setting or involved in psychological therapies. HoNOS is a clinician-rated measure, while CORE-OM is a self-report measure.
3. Most rating scales in psychiatry are completed by clinicians. The patient voice has tended to be ignored in the development of instruments. Potential problems with clinician ratings include the cost, training, potential for gaming, reliability of the assessment and inter-rater reliability. Potential concerns with self-reports from patients include concerns around social desirability, idiosyncratic completion, and use with patients who are acutely unwell. There is usually not much agreement in ratings between clinicians and patients on different scales or even when using the same instrument.
4. International approaches to routine outcome measurement have primarily been driven to inform programs and systems, rather than helping to inform individual patient-level treatment outcomes. Purchaser-driven pressures in the US have led to the mandating of routine outcome assessment in a number of different settings, for example the Veterans Administration and the Ohio Department of Mental Health. Outcome assessment is linked to performance measures on recording compliance and agency certification. Australia has the most coherently developed approach to treatment-level routine outcome assessment. They have mandated the use of HoNOS as a standard outcome measure for all patients, as well as a self-report instrument. Perceptions of the value of the outcome measurement system seem to be mixed. Equal numbers seem to find HoNOS to be of value and not of value, although more positive observations have been made about the user-rated outcome measures.

5. In the UK, an Outcomes Reference Group was established by the Department of Health in 2002 to advise on best practice guidance which culminated in the Fonagy report. One of the key conclusions was that it was essential for local Trusts to develop systems to use local outcomes data effectively to inform service delivery, before the data could be used for higher level purposes (such as productivity measurement). The HoNOS was recommended for use for all patients on enhanced CPA care, with at least one measurement taken per year. If other measures are adopted, they should be able to generate an equivalent HoNOS score. These recommendations were subsequently issued as Technical Guidance for the minimum requirements to complete the mandatory Mental Health Minimum Data Set (MHMDS) for England since April 2003. The Fonagy recommendations therefore become a mandate by the DoH.
6. Alongside the Outcomes Reference Group, pilot work in four mental health Trusts tested the practicalities of using different outcome measures, including HoNOS. Key conclusions from the pilots were that response rates for the clinician-rated measures were much higher than for user-rated measures, outcome measurement relies critically on the availability of additional data such as diagnosis in order to inform treatments, and most notably Trusts lack effective systems to promote feedback to clinicians so that the benefits of outcome measurement can be appreciated.
7. While there is consensus that outcomes should be routinely measured, there is in fact limited evidence that routine outcome measurement can deliver improvements in local service delivery and patient-level care. The evidence from RCTs suggests that one-off or infrequent outcome measurements have very little effect on improving quality of life or other subjective secondary outcome measures. But outcome measurement that is done longitudinally and more regularly (with more than one or two administrations) can significantly improve patient's quality of life or reduce psychiatric admissions.
8. HoNOS was developed by the Royal College of Psychiatrists' Research Unit (CRU) between 1993 and 1995 and is recommended by the National Service Framework (NSF). HoNOS instruments are available for Children and Adolescent services, older people, forensic services, learning disabilities and acquired brain injury. There are 12 items on HoNOS each scored from 0 (no problem) to 4 (severe problem) yielding a total score in the range of 0 (best) to 48 (worst). The HoNOS takes on average between 5 and 15 minutes to complete, depending on the experience of the rater and the complexity of the patient's problems. The rating period covers the previous two weeks. CRU provides training for raters and training for trainers, however the cost is not trivial. HoNOS has undergone a number of independent studies to examine its psychometric properties and has been found to have good validity, and adequate reliability, sensitivity to change and acceptability. Comparing completion rates for HoNOS in the MHMDS in 2004/05 and 2005/06, the coverage of HoNOS dropped from 44 to 37 providers (out of 84), although the overall completion rates for those who submitted HoNOS returns remained around 9.5%. Due to the drop in the number of providers completing HoNOS, the overall coverage dropped from around 5% to 4%. Time-series activity and outcome data exists from 2003 onwards in the MHMDS for use in a productivity index, though data quality suggests only the latest 2 years should be used.
9. CORE-OM was developed between 1995 and 1998 through a multi-disciplinary team of practitioners representing the major psychological therapy professions. CORE Information Management Systems (IMS), a not-for-profit organisation, provides support to the CORE Network of users who can voluntarily anonymously donate aggregate to the National Research Database. CORE-OM is a free tool and comprises 34 items each with 5 levels from 0 (not at all)

to 4 (all the time), giving an average score of between 0 and 4. It takes approximately 5 minutes for the patient to complete. Normative tables exist with cut-offs (severity bands) for clinical and non-clinical samples of the UK population. There are also two software versions, namely CORE-PC and CORE-NET. Studies to examine the psychometric properties of the CORE-OM have found it to have reasonable reliability, validity, sensitivity to change and acceptability. CORE-OM is believed to be the most widely used outcome measure in psychological therapy and counselling services, used in around 250 services in the NHS. The database which CORE IMS holds, covers around 100,000 patients per annum. Around 30 Mental Health Trusts and around 75 PCTs use the software version in psychotherapy and counselling services. Time-series activity and outcome data exists from 1999 onwards in the CORE IMS Database, but access would need to be explored for use in a productivity index. Data quality in early years would need to be tested.

10. Generic outcome measures such as the QALY are often not applicable for mental health patients and are not typically used in routine care. There would be considerable challenges in both trying to value or map either the HoNOS or CORE-OM to a generic measure like a QALY for use in a productivity index. The complexity of the valuation problem is enormous and while mapping could in principle be done, it would require the collection of both sets of data, the disease-specific and generic measure on the same set of patients. Mapping depends on the degree of overlap in content between the instruments being mapped and this will be questionable in these circumstances.
11. In addition to the literature search, interviews were held with 28 policymakers, academics, and NHS staff involved in outcome measurement, including managers, clinicians and commissioners. Ethics approval was obtained. The themes emerging from the interviews included the main barriers to outcome measurement, the incentives that drove the process, the choice of the instrument, and the policy context. Probably the most crucial barrier to the introduction of outcome measures is that clinicians are unable to see the clinical benefits, partly because they have not been given a clear rationale for their use, partly because they are simply told to complete scores, but primarily because they never receive any feedback on them. Many see it as a paper-filling exercise for managers. Unreliable and out-of-date IT systems and a lack of IT skills was another major barrier. For clinicians to get valuable feedback, often required them to design their own IT system. Another one of the main barriers was the lack of an outcome-oriented clinician culture, one which is open and learning rather than fearful. Peer pressure was seen as potentially useful. Time was not seen as a major barrier and nor was training.
12. In terms of incentives that might drive the implementation, clinical champions enthusiastic about outcome measurement, were seen as imperative. Management support was seen as essential, though a top-down push from management was seen as a disincentive. A key theme which emerged is that routine outcome measurement is unlikely to be viewed as a mandatory activity, until it becomes a biting target in the Healthcare Commission's performance management regime. The Healthcare Commission could incentivise uptake of HoNOS by including coverage in the MHMDS as a performance measure. This is currently the case for coverage on the ethnicity completion within the MHMDS, though such a performance target would drive managers rather than clinicians. Application for Foundation Trust status on the back of the Healthcare Commission annual health check was seen as another key driver. Trusts felt they had a strong incentive to show effectiveness. Commissioners were seen as weak and not a strong driver for the uptake of routine outcome measurement. There were mixed views about whether Payment by Results (PbR) would be a potential incentive to drive the introduction of routine outcomes in the future. Some felt that it would incentivise only managers, but not clinicians, though PbR was felt to still be a long way off for mental health. Interestingly, a modified version of HoNOS called HoNOS Plus has been piloted as an assessment tool for determining 13 activity categories for use in PbR.

13. Current policy developments on outcome measurement fall under the National Outcomes Measures project of the Care Services Improvement Partnership (CSIP) who are trying to ensure the implementation of some of the recommendations from the Fonagy report. They are also responsible for the development of a Compendium of outcome measures which will outline the available measures, their psychometric properties, their uses in different circumstances, and the copyright issues. Outcome measurement does not however receive high policy priority, for example in the NSF.
14. There is a large literature about calculating price indices for mental health care in the US which specifically explores the change in the composition of treatments over time. Rising expenditure on mental health has generated considerable interest in constructing price indices, for example for major depression, schizophrenia and bipolar disorder. This literature shows it is important to focus on the direct medical costs of treating an episode of illness, rather than changes in the prices of the inputs used in treatment. For all three disorders, studies suggest that the price of treating an episode have declined in recent years. This is contrary to many of the officially reported figures because the conception of output allows for a substitution among inputs as a result of technological change. These studies cannot readily be reproduced with existing UK data. While output indices could in principle be generated from the MHMDS and there may even be the possibility of tracking compositional changes in treatment over time, there is no way of producing price indices equivalent to those generated from the medical claims data in the US. No costing data is at present readily or routinely available.
15. It is probably premature to incorporate mental health outcomes into a productivity index. Outcome data collection needs to be improved first and recommendations for this include improving policy guidance and IT systems, and ensuring feedback mechanisms, management support and an outcome-oriented culture.

Appendix 3 – Outcome Measures for the Assessment of Treatment Results in Chronic Obstructive Pulmonary Disease

Research Report for the OHE Commission on NHS Productivity

Helen Starkie and Andrew Briggs, University of Glasgow

EXECUTIVE SUMMARY

The aim of this research is to assess those outcome measures that could and should be routinely collected from chronic obstructive pulmonary disease (COPD) patients: the focus being upon the health benefits achieved for patients following treatment in the NHS.

Five discussion areas were proposed:

Are any data available on changes in the mix of treatments in COPD over time? Are there data relating to the costs of the changing mix of treatment?

It is well known that COPD, as a disease, is increasing in prevalence. This is due, in large part, to the emerging consequence of past patterns of smoking. This trend in the prevalence of the disease is expected to continue for the foreseeable future. Data are available on the changing mix of treatments overtime – particularly in relation to new drug therapies aimed at controlling symptoms and reducing the incidence of exacerbations. The development of a number of COPD guidelines have contributed to the changing patterns of care for COPD as the clinical community seeks to define appropriate patterns of care for this challenging patient group. However, the licensing of new products, particularly combination products, is probably responsible for the most important changes in treatment patterns – with 24% of patients now prescribed an ICS/LABA combination therapy compared to just 1% six years ago.

With this changing mix of pharmacotherapy has come a consequent increase in costs for pharmaceuticals. Costs of other forms of care are more difficult to quantify. Hospitalisation costs will increase with increasing prevalence of COPD. Nevertheless, costs of care at a patient's level might be expected to decrease with improved pharmacotherapy; leading to reduced incidence of exacerbations requiring intervention, particularly in secondary care.

What outcome measures are currently used within COPD? Which of these are believed to be the most sound methodologically and which are used most commonly in practice?

A wide variety of potential outcome measures exist within COPD. These include hard endpoint data, clinical measures and health outcome measures. Hard endpoint data, such as mortality, hospitalisations, and possibly exacerbations involving contact with health services, have the potential to be monitored routinely using existing systems.

Clinical measures that have been identified as being appropriate and practical to collect on a routine basis include spirometry, body mass index and possibly exercise tolerance. Many of these measures are already routinely collected. Clinical measures are limited in their scope to address questions of health benefit to patients and may only be weakly correlated to actual health benefit as experienced by the patient. Health outcome questionnaires on the other hand, specifically lend themselves to monitor health benefit as perceived by the patient.

In lieu of a gold standard to assess the methodological soundness of a HRQoL instrument, standard methods can be utilised which assess important properties of these questionnaires, including reliability, validity, responsiveness and precision. The evidence of a recent structured review of health outcome measures was reported.

Of the disease specific measures, the Chronic Respiratory Questionnaire (interview administered) and the St George's Questionnaire have been widely used in COPD populations and the bulk of evidence shows that these two measures hold the four desirable psychometric properties of reliability, validity, responsiveness and precision.

For the generic measures, the SF-36 is by far the most widely used (utilised in 90 studies) out of all the generic measures and correspondingly there is a wealth of evidence to support its methodological qualities. Comparing across all disease areas, similar numbers of studies used the EQ-5D as the SF-12 and these two measures both showed clear fulfilment of the ideal psychometric properties and have been validated in similar number of studies. All three of these measures have the advantage that algorithms exist to allow the generation of utility preference data suitable for estimating QALYs.

What are the practical issues about collecting these outcome measures routinely? How difficult would it be to collect these measures/data within the NHS? Are clinicians likely to use them?

Five outcome measures that showed sound methodological properties were identified. However, for use within a routine setting an outcome measure needs not only to be methodologically sound, its collection also needs to be acceptable to the patient and clinician in the context of a busy clinic.

The COPD population tends to be elderly and some have low educational standards. In a busy clinic or hospital, there may be little support available for a slow or poorly educated patient who may struggle with reading and/or understanding the questionnaire. Each patient must be able to fill out the questionnaire unsupported, or with only minimal assistance. The COPD population is associated with a range of co morbidities so, not only must the measure be sensitive to small changes in the disease; it should be able to take account of co-morbidities. COPD patient characteristics have important ramifications for the appropriateness of any questionnaire used as well as the administration of that measure.

Feasibility of routine collection and how an instrument might best be administered was also considered. A series of questions that would need considering prior to the implementation of routine data collection was presented. Potential barriers and solutions to the acceptance of routine outcome measurement by medical staff were highlighted. Overall, the need for a measure that is simple and quick to administer in the context of a busy clinic, but which is capable of generating a utility based preference score suitable for calculating a QALY suggests either the EQ-5D or the SF-12 as the most appropriate instrument.

Can these disease specific outcome measures be converted into a QALY or some other generic measure? How easily can the disease specific measures be mapped onto the generic measure? Could these outcome measures be used to monitor changes in productivity?

Examples were given of cases where the disease specific measure, the SGRQ, was mapped to the EQ-5D. In addition, the creation of a utility measure, the SF-6D, for both the SF-36 and the SF-12 was described. Monitoring productivity within the NHS may be achieved through incorporating HRQoL values into a national audit, such as the National COPD audit. Direct monitoring of a utility measure is preferable for the ultimate aim of generating a QALY. Nevertheless, indirect monitoring would be possible through the use of a mapping between the disease specific instruments and utility measures. Finally, consideration should be given to whether it is possible to generate an estimate of

QALYs associated with COPD by a simple approach of ascribing utility decrements to observed COPD hospitalisations and exacerbations combined with background disease severity. Such an approach would be easy to implement and would not require additional data collection from patients.

Is there a time series of these outcome measures available? Going back how many years? If not, were there measures used previously in COPD and could these be mapped to the current outcome measure to construct a time series?

A number of databases exist within the UK, both in the primary and secondary sector as well as general health surveys and a national audit for COPD. Within these databases, physiological data is routinely collected, specifically, acute exacerbations (resulting in hospitalisations), mortality and lung function scores. Outside of clinical trials and observational trials, routine collection of HRQoL data is not collected. The only way to construct a time series would be to use a simplified mapping process based on linking utility decrements to hospitalisation and exacerbations.

Appendix 4 – Consideration of Routine Measurement of Colorectal Treatment Outcomes

Research Report for the OHE Commission on NHS Productivity

Paul Trueman, Matthew Bending, Steven Duffy and Paul Kind,
York Health Economics Consortium

1. BACKGROUND

Attempts to capture the outputs of the National Health Service to date have focussed on measuring the volume of health service activities, such as hospitalisations or surgical episodes. Whilst such information provides a measure of activity, it makes no attempt to capture the primary output of the health service, namely improved health outcomes.

The Office of Health Economics (OHE) Commission on NHS Productivity commissioned the York Health Economics Consortium (YHEC) to undertake research to address the following points:

- Identify information that is collected routinely on patients with colorectal cancer treated by the NHS;
- Identify their relevance to the measurement of NHS productivity;
- Provide recommendations on whether patient outcomes could be routinely incorporated into data collection on patients treated for colorectal cancer.
- The research builds on the findings of previous research conducted jointly by YHEC and the University of Sheffield.

2. METHODS

The research comprised three distinct tasks:

1. A review of the information routinely collected on patients with colorectal cancer, treated in the NHS;
2. A literature review of published evidence on outcomes measurement in colorectal cancer;
3. Interviews with stakeholders to identify any unpublished information or ongoing attempts to capture patient outcomes routinely in practice.

Finally, a framework was developed to determine the most appropriate method for routinely capturing patient outcomes in practice, which considered the feasibility of collecting data, the burden to patients and health service professionals and the ability to use the outcome as a measure of health service productivity.

3. FINDINGS

The findings suggest that routinely collected data on patients with colorectal cancer focus predominantly on hard, clinical outcomes. These outcomes include mortality rates, survival rates, intervention rates and adverse event rates (for example, adverse events associated with surgery). These measures are all well-defined and routinely captured in cancer registries. No large-scale attempts to routinely capture patient reported outcomes or quality of life were identified.

Whilst these outcome measures fail to take into account patient outcomes they are highly relevant to the management of colorectal cancer. Mortality and survival remain critically important outcomes for a significant proportion of patients with colorectal cancer. As such, efforts to capture this data, for example in cancer registries, should be continued and improvements in survival are a fair reflection of improvements in health service productivity in this indication. However, additional

information on quality of life would be a valuable adjunct to this information, particularly in palliative care patients, where one of the main objectives of treatment is improved patient quality of life.

The literature review identified the most widely used patient reported outcome and quality of life measures that have been used in studies of colorectal cancer. The studies included in the literature review were randomised and observational studies of radiotherapy, chemotherapy and surgical interventions for colorectal cancer. The most widely used patient reported outcome measures were those developed by the European Organisation for Research and Treatment of Cancer (EORTC), including the EORTC-QLQ C30 for use in all cancers and the EORTC-QLQ CR38, developed specifically for use in colorectal cancer. A second cancer specific outcome measure, the Functional Assessment of Cancer Therapy (FACT) questionnaire, was also widely used. In addition to the cancer specific outcome measures, two widely used generic patient outcome measures were also identified. The Short-Form 36 (SF-36) and the EuroQol EQ-5D were identified as commonly used measures in study settings.

Interviews with stakeholders identified some local attempts to capture patient outcomes data in cancer patients, including colorectal cancer patients. These studies have examined the feasibility of collecting such data, including the use of computer interfaces to reduce the burden of data collation. The studies have shown a high degree of acceptance by patients although adherence tends to fall over time, suggesting that frequent, repeated applications of patient reported outcome measures may create a burden to patients, resulting in incomplete data.

4. RECOMMENDATIONS

Based on the above findings, a number of recommendations for consideration by the OHE Commission have been developed:

- Measuring mortality and survival rates remains critical to the assessment of productivity in colorectal cancer care and should continue to be routinely collected in cancer registries;
- Routine data on patient reported outcomes would help to provide a valuable adjunct to clinical outcomes data;
- The use of generic scales, such as the SF-36 or EQ-5D is the preferred approach to capturing routine information on patient reported outcomes, as it allows for comparison with other conditions and has the most potential to be used in the assessment of health services productivity;
- Patients should be asked to complete quality of life surveys at infrequent intervals (e.g. diagnosis, 1 year follow-up, 5 year follow-up) and ideally during a routine consultation;
- In addition to this, transitional change questions should be routinely incorporated into each scheduled visit to determine whether patients consider their health to be improving over time;
- The value of the information collected in routine practice should be regularly reviewed to ensure that it contributes to the measurement of productivity and is acceptable to patients and healthcare professionals.

Appendix 5 – The Measurement of Responsiveness

Research Report for the OHE Commission on NHS Productivity

Nigel Rice and Silvana Robane, University of York

SUMMARY

The concept of health system responsiveness

Health system responsiveness is a concept promoted by the WHO and introduced with the release of the World Health Report. Together with health and fairness of financial contribution, the report defines responsiveness as an intrinsic goal of health system performance. In broad terms, health system responsiveness has been defined as the way in which individuals are treated and the environment in which they are treated and importantly, encompasses the notion of an individual's experience of contact with the health system. These experiences are measured along a number of different domains which can be classified into two broad dimensions, namely respect for human rights and client orientation. Although health and responsiveness are viewed as independent intrinsic goals of a health system, they are not entirely unrelated. A responsive health system will contribute indirectly to improved health outcomes, and is assumed to contribute directly to patient welfare.

The measurement of health care quality covers a multitude of outcomes and processes and has been described as including the dimensions of technical quality, process quality and structural quality. The concept of responsiveness relates to this literature through interpersonal aspects of process and structural quality, which include the quality of the interaction between the patient and the provider. While there is an overlap between responsiveness and aspects of quality of care, the literature on the latter does not seem to encompass all dimensions relevant to a responsive system. Responsiveness may be more related to patient satisfaction. Satisfaction is a subjective measure which can be influenced by expectation, whereas responsiveness attempts to reflect actual experiences of contact with the health system distinct from expectations.

The measurement of responsiveness

The development of a WHO instrument for measuring responsiveness was based on an extensive literature review of health system quality and patient satisfaction. The results of this informed the selection of the domains of responsiveness that appeared to be valued most by individuals in their contact with health systems. The selection process also made referred to pieces of work by other organisations such as the Agency for Health Research and Quality. Based on this research, eight domains were settled upon:

- Autonomy
- Choice
- Clarity of communication
- Confidentiality of personal information
- Dignity
- Prompt attention
- Quality of basic amenities
- Access to family and community supports

An important development within this process was a move away from a reliance on key informants to elicit views on health system responsiveness towards a household survey of respondents. Many of the domains are orientated towards the perspective of the patient rather than a health professional. The measurement of responsiveness is obtained by asking respondents to rate their most recent experience of contact with the health system within each of the eight domains. Each domain is measured using a categorical variable.

The most ambitious attempt to date to measure and compare health system responsiveness is the World Health Survey (WHS) – an initiative launched by the WHO in 2001 aimed at empowering countries to monitor health and health care outcomes through the fielding of a comprehensive, reliable and comparable survey instrument. Seventy countries participated in the WHS 2002-2003, and data collection was on a modular basis covering information on health insurance, health expenditures, socio-demographics and income, health state valuations, health system responsiveness and health system goals. The survey is divided into two sections - the first a household questionnaire and the second section is an individual questionnaire, however there has not been any formal analysis of the WHS responsiveness data.

Practical issues in the measurement and analysis of responsiveness

The concept of responsiveness relates to all contact with the health system and while this clearly encompasses health services and health care, it also, in principle, relates to less easily defined areas of health system activity such as health promotion. Measuring responsiveness for contacts with public health initiatives and messages is somewhat more difficult.

Another potential problem that arises in the measurement of health system responsiveness when individuals are for whatever reason denied access to services. A system that restricts access to certain individual and focuses attention and resources on those provided with care can potentially achieve a higher responsiveness rating than a system operating with an equal level of resources and that treats all.

The multidimensional nature of responsiveness raises a number of concerns around how best to combine responses to each domain to construct an overall composite measure to allow health system responsiveness to be summarised and compared. One suggested approach is to weight responses on each domain as rated by respondents. Such weights can be derived from household surveys where alongside questions about actual contact with the health system, respondents are asked to rate the relative importance of the domains.

The use of a common set of weights to compute an overall measure of responsiveness is helpful when undertaking cross-country analyses to ensure greater comparability of the survey instrument. However, there is a concern that the application of a common set of weights might prejudice certain populations in favour of others – weights therefore may vary across countries due to a variety of cultural and political factors. The challenge of how to appropriately compare across institutional settings and populations is a central feature of comparative work for all public services. A fundamental problem is that studies aimed at comparative inference have rarely taken into consideration possible variations in cultural expectations that might impact on reporting behaviour. To this end, effort has been placed in producing more objective measures of responsiveness and developing instruments that are relevant across cultural settings.

Recently methods of anchoring vignettes have been promoted as a means of controlling for systematic differences in preferences and norms when responding to survey questions. Because individuals are asked to evaluate these hypothetical cases in the same way as they evaluate their own experience of the health care system, responses to the vignettes allow the investigator to model the response scales as a function of the characteristics of respondents.

Goal attainment as defined by the WHO relates to both the level and distribution of responsiveness. The majority of the literature on responsiveness has, however, focused on the former, but the distribution of responsiveness and how this varies across socio-economic group is also relevant to policy-makers.

Conclusions

Both health and non-health benefits have been promoted as appropriate outcomes on which health systems can be judged. The concept of responsiveness, which can be seen as an elaboration of the better known concepts of health care quality and patient satisfaction, has been promoted by the World Health Organisation as a credible method of measuring non-health benefits and how these impact directly on patient well-being. However there is little empirical work on the use of this instrument thus far.

With regards to the use of such an instrument within the context of performance management in the National Health Service much could be gained by focusing on elements of responsiveness available in current survey questionnaires, particularly domains that appear to be valued most by patient. It would further be fruitful to cross-reference the concept of responsiveness to the National Patient Experience Survey to assess the degree to which this survey could be used to inform measures of responsiveness.

Appendix 6 – Broader Societal Costs and Benefits

Research Report for the OHE Commission on NHS Productivity

Albert Chibi and Clive Pritchard, Office of Health Economics

EXECUTIVE SUMMARY

We conducted a literature review of the economic evaluations literature to identify studies which had considered wider social costs beyond the health care costs to which the remit of health technology assessments, including those by the National Institute for Health and Clinical Excellence (NICE), are often restricted. Studies we considered relevant for our review were original cost-effectiveness and cost utility analyses published from 2000 which had been undertaken in a UK context and had assessed the impact of illness and treatment on work time (paid or unpaid), leisure time, informal care or schooling.

The typical focus of the 50 studies found to be relevant from the Health Economic Evaluations Database (HEED) and the NHS Economic Evaluation Database (NHS EED) was time lost from paid work (46 studies). Only 17 studies explicitly stated the approach taken to the assessment of wider social costs, of which 16 used the human capital approach, one used the friction cost approach alone and one used both. No study explicitly employed the US Panel approach although, according to this approach, some studies would be regarded as implicitly allowing for the impact on time by means of the QALY measure. As far as therapeutic areas are concerned, the most commonly studied chapter of the ICD-10 classification system was that covering mental and behavioural disorders, with depression being studied more than any other condition.

Randomized controlled trials (RCTs) were the most common basis for the studies we reviewed. Given the need for self-reporting when making allowance for the time of patients and others, RCTs present a particularly amenable vehicle for evaluating wider social costs. However, the short duration of RCTs may give an insufficient time horizon, for subscribers to the human capital approach, to take account of all relevant costs. In our sample, 22 RCT-based studies (of the 26 for which it was stated) had a time horizon of a year or less. The only studies with a time horizon extending to the whole lifetime used a modelling approach. Some studies which declared that the human capital approach was being used also referred to the friction cost approach or the distinction between short term and long term impact. However, the restricted time horizon would tend to mitigate any potential difference between the results obtained under the two approaches.

One study which employed the friction cost approach as well as several variants on the human capital approach illustrated the potential impact of employing different assumptions, with cost per QALY estimates straddling NICE's benchmark of £30,000 per QALY. However, a strict interpretation of the human capital approach was found to give similar results to the friction cost approach. This study shows that not only are there disagreements about the principles involved, but also that differences in interpretation are possible when attempting to implement one general approach.

Although we did not investigate whether, across the whole group of studies, the inclusion of wider social costs made a difference to the cost-effectiveness results, we undertook a more detailed inspection of the studies in depression, this condition appearing with the greatest frequency in the sample. All seven studies concentrated on sickness absence from work. Four were based on RCTs and three on a modelling approach, but none used a time horizon in excess of 12 months duration. In only one study did inclusion of productivity costs appear to make an appreciable difference to the results, accounting for an overall cost saving in the post-treatment period under cognitive

behavioural therapy compared with usual care. Apart from the short time horizon (potentially important for a recurrent condition such as depression), an issue with RCTs is the extent to which they are powered to detect differences in costs, including productivity and other time costs. Greater use could potentially be made, within the clinical trial setting and in routine clinical practice, of the range of instruments developed for measuring resource use to capture wider social costs associated with illness and treatment.

Appendix 7 – Discrete Choice Literature Review

Emma McIntosh, University of Oxford

Methods

A systematic review of discrete choice experiments (DCEs) in health care was carried out to establish what health and non-health attributes matter to patients. The systematic review explored all discrete choice studies carried out in health care both in the UK and internationally. The literature review used the accessible data sets available to the University of Oxford search engines (OVID and Silverplatter), including the following: CINAHL, EMBASE, MEDLINE, JSTOR, AMED and ECONLIT. Further, in order to identify any unpublished work carried out by authors specialising in this field (both in the UK and internationally) authors were also contacted directly by e.mail. The key search terms used were as follows: discrete choice experiment\$; discrete choice conjoint experiment\$; conjoint measurement; conjoint analysis; discrete choice model\$; stated preference discrete choice experiment\$; discrete choice conjoint experiment\$; paired comparisons; binary choice; stated preference; conjoint; functional measurement; and pairwise choice\$.

Upon completion of the systematic literature review and collection of articles a proforma was devised to identify and collect the relevant information from the articles. This proforma collated information on the following: Place of study (UK or international); Type of study (health system, pharmaceutical, methodological); Respondent type (patients, general public); Type of attributes identified in study (health, non-health, process); Quality of study (using design criteria) and finally the statistical significance of attributes. Using this proforma the data were categorised accordingly and a discussion drawn up as to the importance of attributes to patients and the general population more generally.

Results

Using the search strategy outlined above the following results were obtained: Medline (1950-week 2 Sept 2007) 2,309; EMBASE (1980-2007 week 38) 1,885; CINAHL (1982-September week 2, 2007) 1,743; JSTOR 36,258; ECONLIT (1969-1999, 2000-2005, 2006-2007) 1,499; AMED (1985-Sept 2007) 66; Journals at OVID 7,391. Abstracts were read for relevance to the health care setting, duplicates removed, methodological non-applied papers excluded, non-patient and non-population preferences were excluded and the final number of relevant articles was 145. The relevant papers cover a vast spectrum of health and non-health attributes from the methodologies of both discrete choice experiments and conjoint analysis. The majority of studies have been carried out, or have elicited UK patient samples, in the UK (52) USA (29) and Australia (14) followed by a number of other European countries such as Spain, the Netherlands and Germany. The majority of studies using discrete choice experiment methodology reported statistical significance of individual attributes whereas many of the conjoint analysis studies reported only utility scores or relative importance of attributes in percentage terms.

No attempt was made to grade the quality of the papers in terms of design of experiment, criteria used for design, properties of design nor statistical technique used. The vast majority of papers, even where health was included as an attribute, incorporated non-health and process attributes of some form. A large number of the papers also estimated willingness to pay values for such attributes. There is a vast breadth of health, non-health and process attributes, covering a huge spectrum of disease areas – many of which are significantly valued by patients and the population.

Importance of attributes to patients

By categorising the 145 papers according to whether the majority of attributes were associated with 'Health', 'Process of care', 'Pharmaceutical', 'Health care system' or 'Other' the majority of studies were concerned with 'Health' attributes (n=88) this was followed by 'Process' attributes (n=44), 'Health care system' attributes (n=5), 'Pharmaceutical' attributes (n=5) and 'Other' attributes (n=3). From this it would seem that the majority of studies in the DCE/Conjoint field are being driven by a desire to identify the statistical significance, value of and trade-offs between health attributes. Indeed just under half of these 'Health' studies included cost as an attribute in a bid to measure respondent willingness to pay (WTP) for health care attributes (n=42). This large proportion of 'valuation' studies embedded within the DCE/conjoint studies reflects the academic trend of the last decade to obtain a common numeraire with which to be able to compare attributes. Of those studies mainly interested in identifying and measuring 'Process' attributes, just under one third of studies (n=13) estimated WTP values.

Without having specific information on the research purpose behind individual studies, nor the aims of the research commissioner or individual researcher it is difficult to say exactly what is driving the research on the attributes identified in this literature review. As outlined above, the range of attributes is huge although it is clear that the majority of commissioned research in this area still seems to be concentrated on measuring and valuing 'Health' related attributes. While 44 studies were concerned mainly with 'Process' attributes it should be noted that a large number of the 'Health' classified articles also contained process attributes. What is clear from the 'Health' studies is that patients place significant value on 'process' attributes even when they are included alongside health care attributes. This is evidenced in the majority of 'Health' classified studies.

Without being able to judge fully the quality of each and every study as well as the research context it is challenging to draw conclusions regarding the extent to which health care providers, such as the NHS, should be measuring and prioritising non-health, process attributes. It can be seen from this literature review that the measurement and valuation of non-health process attributes are certainly a large and emerging field of research. Based on those process attributes appearing as statistically significant in this review it could be surmised that in the situation where all else is equal between two health care interventions then that intervention which contains process attributes of the types identified in this review should certainly be recommended as the optimal intervention. The extent to which process attributes should be given priority over health care attributes is somewhat more complex however. Where individual studies identify this to be the case it is recommended that much more detailed analysis of the quality, funding and context of such studies should be carried out.

In the studies where the majority of attributes were mainly 'Process' it may be the case that re-organisation of the service in terms of these attributes, i.e. location, staff type, facilities, mode of administration, information available for example may in fact be able to be carried out at neutral or minimal cost. In such cases it is a recommendation that researchers involved in such studies go the extra step of exploring the cost-effectiveness or cost-benefit of such re-organisation to aid implementation of research.

A further complexity is where process attributes are highly correlated with attributes such as 'waiting time'. Whilst 'waiting time' is a process-type attribute it is certainly the case that longer waits will be associated with poorer health in many situations. One recurring process attribute identified in the review as highly valued by patients is related to choice of, continuity with and type of medical staff seen e.g. GP. This is a strong thread through many of the 'Process' studies identified in the literature review. The costs or re-organising services in line with such skill-mix preferences is possibly one area which could be explored in more detail.

Appendix 8 – Review of Methods for Mapping between Condition Specific Measures onto Generic Measures of Health

Research Report for the OHE Commission on NHS Productivity

John Brazier, Yaling Yang and Aki Tsuchiya, School of Health and Related Research, University of Sheffield

Executive Summary

A common approach to assessing the outcomes of health care is to obtain patient reported descriptions of health status across various dimensions and then to apply a standardised numerical scoring system. There are many different measures of health, including several hundred condition specific measures of health designed for use in specific medical conditions or groups of condition (Spilker et al, 1990). There are generic measures designed to cover the core dimensions of health that are relevant across all medical conditions.

Preference-based measures of health have been designed to generate the health state utility values required to calculate QALYs for assessing the cost effectiveness of interventions. These are usually based on generic instruments that permit comparisons between patient groups (e.g. EQ-5D). Even for assessing clinical effectiveness, it could be argued that a preference-based index is necessary to deal with trade-offs made between outcomes.

There has been a lively debate amongst health economists about the most appropriate generic measure. Different preference-based measures have been shown to generate different values on the same sample of patients. Furthermore, many key clinical trials on the efficacy of new interventions do not have a generic measure. This presents a barrier to populating economic models with the best evidence on effectiveness.

One solution to the problem of having different measures of health has been to try to map between measures using judgement or statistical inference. There are two approaches to mapping. One uses judgement to undertake the mapping (e.g. using panels of experts) and the other is done empirically using a data set with the two measures administered on the same patients.

Mapping using judgement

In this approach, components of a non-preference based measure of health or quality of life would be assigned by judgement to specific domains and levels of a generic preference-based measure (such as the EQ-5D).

The main criticism for this approach is its arbitrariness. Furthermore, it does not involve any attempt to estimate the uncertainty around the mapping. The validity of the mappings is questionable and the only way to test the validity of the mappings is by empirically comparing the judgements against real data. Ultimately a much better approach is to estimate the relationship between the measures empirically by statistical inference.

Empirical mapping

The approach of empirically mapping a health measure onto a (usually) generic preference-based measure is also known as 'cross walking' between instruments. It requires the two measures to be administered on the same population. Having obtained the data there are a number of choices regarding the possible specification, the statistical technique for estimating the mapping function and how performance should be assessed. These are now discussed in turn.

There are different possible specifications for the mapping function. The simplest additive model is to regress the target measure onto the total score of the starting measure. A more sophisticated approach to modelling the relationship between measures is to regress separate models for each dimension of the target instrument.

Estimation and model performance

Most mapping functions have been estimated by OLS, though some researchers have explored Generalised Linear Models with random effects, Adjusted Least Square Regression Model (ALS), Tobit Model, Censored Least Absolute Deviation Model (CLAD) and non-linear models.

The performance of models can be assessed in a number of ways in the literature. It is common to report explanatory power in terms of adjusted R-Squared, but this has limited value for comparing models estimated using different methods of estimation.

Ultimately the purpose of modelling these data is to predict values in other data sets. One way to evaluate models has been to examine the difference between predicted and observed values at either the aggregate level by calculating Mean Error (ME); or the individual level by calculating the Mean Absolute Error (MAE) or the Root Mean Squared Error (RMSE).

Literature review

The review aimed to address the following questions:

- What instruments have been used in the mapping studies?
- What medical conditions/diseases are covered by existing mapping studies?
- What methods have been employed to undertake the mappings?
- How have the mapping functions been assessed in terms of statistical performance?

Findings

Description of studies

A total of 28 papers were identified for the review. (These 28 papers covered 119 different estimation models). Mapping between HRQoL measures is a new research area with most papers (26 out of 28) published or produced after 2000, with the remaining 2 papers published in 1997 and 1998.

Data sets

Twenty-seven out of the 28 studies involved the mapping of non-preference based measures of health (the "starting measure") onto preference-based measures of health (the "target measure"); the exception being between preference-based. The most popular target measure used for mapping was the EQ-5D with 16 studies. On the right hand side of the mapping equation, the most widely used starting measures were the related generic health measures of SF-12 (n=7) and SF-36 (n=5), and the remainder consisted of various condition specific HRQoL instruments covering asthma, overactive-bladder, obesity, cancer and heart disease. Clinical trials were the most common source of data.

Specifications

Most studies used the index generated by the preference-based measure as the dependent variable and estimated models using OLS. Some explored Maximum Likelihood, Adjusted Least Square Regression Model (ALS), Tobit Model, Censored Least Absolute Deviation Model (CLAD) and non-linear models. Only four studies had dimension scores as the target measure and in all cases these were the 5 dimensions of the EQ-5D.

Performance

Overall, models mapping a generic onto a generic preference-based measure (e.g. SF-12/36 to EQ-5D, NHP to SF-6D, and SF-36 to QWB) achieve R² or adjusted R² terms of more than 0.5 within sample. The fit of mappings functions from condition specific to generic measures is more variable.

A number of studies have noted that the degree of error is not evenly distributed across the scale of the dependent variable. This problem was shown using condition specific and generic measures as the start measures. Overall, the level of error is far greater at the lower end.

Discussion

All mappings to date have been onto preference-based measures, with a view to assist in estimating QALYs for use in economic evaluations conducted alongside trials or decision analytic models. This reflects a very pragmatic need for health state values for this purpose. This may limit its value for other purposes. Nonetheless there are important lessons to be drawn from the literature.

The question of whether mapping is a viable approach is difficult to gauge without having clear cut criteria for assessing adequate performance. In the ex post situation of the analyst not having any other data, some of the poorer models might be acceptable. This may happen in economic evaluations alongside clinical trials. The concern relates to the degree of error that is acceptable. At the individual level, the mean absolute error was often quite high and larger than published minimally important differences for these measures. What is important is the distribution of that error by severity and in those studies that have actually examined this, the findings seem to suggest that these models tend to over predict at the lower end and under predict at the upper end.

For ex ante, prospective decisions about the instruments the decision maker needs an estimate of the error relevant to the population of interest to decide whether the error in these models is acceptable.

For mapping from conditions specific measures the degree of error tended to be larger, though it varies across patient groups. However, the use of mapping to derive a preference-based generic value from condition specific measures raises a more fundamental concern. Mapping assumes that the preference-based target measure covers all important aspects of health of the non-preference-based start measure. In other words, the strength of the mapping function depends on the degree of overlap between the systems. Where there are important dimensions of one instrument not covered by the other, then this may undermine the model. Even the mapping of two generic measures provides examples of this problem. EQ-5D does not, for example, contain a dimension for energy or vitality. So it is not surprising that in published mapping functions from any of the SF instruments to EQ-5D, energy has a small and non-significant coefficient. Another source of weakness can arise from differences in the severity range covered for given health dimension.

Conclusion

This review addresses a number of questions about the use of mapping between condition specific measures and generic measures. It found a surprisingly large body of literature. The performance of the mappings functions in terms of goodness of fit and prediction was variable and so it is not possible to generalise across instruments. Performance is related to the degree of overlap in content between the instruments being mapped. The current literature is also limited in the way these models have been tested, since most testing has focused on their use at the individual level and yet most uses of these mappings functions are likely to be on subgroups of patients (such as arms of trials or diagnostic group in hospital). Further work is required to test the accuracy of these functions in more relevant contexts and over a larger range of instruments. The use of mapping functions is always a second best solution to using the generic measure in the first place (or arguably using preference-weighted condition specific measure), but it is often necessary for pragmatic ex post reasons and so this remains an important area of research.

Appendix 9 – Using Heterogeneous Metrics to Measure NHS Productivity

Summary of David Parkin's Report, City University, London

Defining the unit of output

The problem of measuring the output of the National Health Service arises from two key elements of the nature of health care:

- it is largely a customised rather than standardised product, with a complex production process. Measuring the volume of output requires a unit of output, and this can be difficult to define if the goods produced by a sector are largely non-standard;
- it is mainly provided publicly rather than sold in a market. Where services are not sold there is no natural unit of output defined by final consumption.

In existing practice, the dominant method of measuring health care output is to define it according to what may be called the *production characteristics* of a good or service as opposed to its consumption characteristics. Units of output are defined by similarities in how much they cost to produce, exemplified by Healthcare Resource Groups (HRGs) in hospital care. However, although a particular HRG may be homogeneous in terms of resource use and the value of those resources, it will only be by chance that it is also homogeneous in terms of its value to consumers. Despite these problems HRGs are recommended as the basic building blocks that should be used for measuring productivity. They are well-defined, easily captured and well-known in the NHS.

Defining the value of output

The various products that make up health care have a complex set of characteristics. For a practical approach it is useful to define these as *primary* and *secondary* characteristics. The primary characteristics are the reasons why the good is desired at all; secondary characteristics are those that affect the enjoyment of the good in the pursuit of the primary characteristics.

The dominant primary characteristic is the desire for an improvement in the person's health over the counterfactual of no health care. It is therefore most important in developing consumption relevant output measures to concentrate on measuring these health changes to define value.

Secondary characteristics are mainly related to how health care is provided – patients might also want the health care process to respect their dignity – rather than its impact on health, or whatever the primary characteristic is. Such secondary characteristics do not define separate health care products, rather they represent aspects of products that are not captured by measuring only the impact on the product's primary characteristics. They should therefore be regarded as *modifiers* of the value of output rather than a separate addition to the overall value of output. They can be often be measured by indicators such as patient satisfaction and responsiveness.

It is not obvious how in practice to include external benefits of consumption, such as benefits to relatives of patients or improvements in workplace productivity. Although these are clearly real benefits, they are the result of health improvements to patients and therefore represent a weighting to health status rather than a separate product. If output is meant to represent only consumer valuations, some of these benefits may be regarded as being internalised by patients as consumers. This would need valuations attached to changes in health status that *explicitly* include such

considerations, which is not the case for existing value sets. It might be possible to add to this an estimate of non-internalised benefits using estimates of the external impact of health state improvements. However, such weightings would implicitly attach a lower weight to those without relatives and those who are not employed. A decision to include this involves value judgements, in this case about equity, as well as technical considerations.

Measuring the impact of health care on health

It is less clear how to measure the outcomes of prevention programmes, both primary and secondary, and public health programmes.

One problem is that it is hard to capture health improvement data from a general or patient population that is not in contact with the health system. For example, a smoking campaign might result in someone giving up smoking and as a result not having a smoking-related illness, but that person would therefore not contact the health system, so it would not be possible to attribute or to record their health improvement.

A second difficulty is that the health improvements are most likely to be reflected in future rather than current output. To differing extents, all health care produces a stream of health improvements that occur in the future as well as the present, but prevention programmes' benefits might *only* be realised many years in the future. It is plausible that population modelling of the impact of these activities on population health over time could provide reasonable estimates of future health improvements.

Valuing improvements using condition-specific instruments

Using a generic health instrument applicable to all types of health care would be the preferred option for measuring NHS outcomes. Different condition specific instruments result in different units of output. There are also conceptual issues. Generic instruments are based on a concept of what health is: wherever they are applied they mean the same thing. Condition specific instruments are based on empirical observation of the relevant condition and are therefore unlikely to be consistent with each other between one condition and another. If it is possible to obtain values for health states from condition specific instruments, either directly or via conversion to a generic instrument, then QALY gains may still be calculated. However, if this is not the case then the condition-specific measures will have different units and another method of aggregating them must be used.

There are operations research techniques that are able to combine output indicators that have different and incommensurable units into an overall estimate of output, such as Data Envelopment Analysis (DEA). This technique relies on various assumptions about the relationship of resource inputs to outputs and uses the concept of best practice to assess the efficiency of different 'production units'. However, it is not possible to say if this is in practice feasible without further enquiry – there is no evidence about using such techniques for this purpose or about using these kinds of data in them. This is certainly worth exploring in future research.

Aggregating different characteristics

For some aspects of public health programmes, it might – as discussed above – be possible to estimate QALY gains, but it is not so clear what units would be used for improvements in, for example, information about health, unless they are estimated by willingness to pay. If we do need to take account of such attributes in looking at overall NHS output, then all types of health output will need to be combined in some way.

One approach would be for the outputs of the different health care products to be combined using explicit weights obtained using a survey. This might take the form of a discrete choice experiment, which would require respondents to make a trade-off between the different primary characteristics that define the product; for example, how much is improved information about health valued relative to improved health itself? The index of output that would be produced could be compared with the overall level of resources that is devoted to the NHS, and productivity change assessed by looking at changes over time.

The issue involved in combining indicators relating to primary and secondary characteristics is similar. Relative weights for primary and secondary characteristics could be established via surveys. This might meet with a problem of lexicographic preferences, where people are unwilling to consider reducing the amount of health produced simply in order to have a better experience of health care. One way to avoid this would be to use a stated preference study that separately obtained values using some common numéraire, for example money.

Conclusions

To summarise:

- For most types of health care, it would be feasible and appropriate to create a QALY type measure of health output using routinely collected data. It would require considerable development however to create algorithms that would translate patient data into QALY gains and require input from population or patient surveys.
- The QALY gain measure would be best calculated using a generic health state measurement instrument, but it would be possible to use transformed data from condition specific instruments applied to some or all conditions.
- It is less clear what can be achieved if not all conditions are covered or if condition-specific data are not transformed into generic form. Operations research techniques such as data envelopment analysis offer a possible means to overcome this problem of incommensurable benefits, but they would require much developmental research to assess their feasibility.
- Output should be organised via HRGs, to facilitate comparison of resource use and output at a micro level and to establish a link between analysis at Trust, PCT and overall NHS levels.
- Indicators relating to the process of health care rather than its outcomes should be regarded as modifying the value of output defined by outcome, not as a separate type of output. Surveys might be used to establish the relative weights of outcome and process factors.
- Health care services whose primary aim is not to improve health should be regarded as a separate output, not converted to QALY gains. Their output might be assessed according to activity. If an overall view is required then again surveys might be used to establish relative weights for combining with other types of outcomes.

Appendix 10 – Undertaking PROMs Collection – Practicalities at the Coalface

Summary of Points Arising During Joint OHE Commission Workshop with BUPA, CHKS and LSHTM

16th July 2007

Purpose

In essence, the aim was to explore from a practical point of view what we are trying to measure by using PROMs (Patient-Reported Outcome Measures) and why; can it actually be done; what are the issues to be raised; and what can one do with the information? Can one actually use it in a practical way to understand how clinical services are performing and indeed how they can be improved?

The OHE Commission brought together three organisations with experience in the collection of PROMs data within the health care sector – BUPA, CHKS and LSHTM – to debate the issues and how the agenda could be taken forward. The day was split into four sessions, as follows.

1. What are PROMs and how are they being used?

PROMs are subjective measures, which are complementary to clinical measures of impairment. Patients may have difficulty in understanding the relevance of, or in interpreting, surrogate clinical measures and are more likely to value measurement of the effect of a disease or treatment on their quality of life. If patients feel more informed by PROMs than by the traditional clinical measures then regular collection of PROMs within the NHS would be a positive move. This information will present a basis on which decisions around choice of treatments/providers can be made. There is a tension between generic measures of utility and the disease specific measures that clinicians favour. A way forward may be to collect generic measures alongside disease-specific ones as a sample and this has actually been advocated by LSHTM in their work on elective surgery.

There are a number of validated internationally recognised tools across a number of different procedures and conditions. However, much of PROMs work within UK health care settings has focussed on elective surgery thus far. BUPA, CHKS and LSHTM have used PROMs to varying extents and have presented such data in various ways. BUPA has published PROMs data on its website. These are the three main examples of PROMs being used at the moment, although there are ongoing national audits and there is a large data gathering exercise through the cancer registries being undertaken.

In general, there is support from the centre in use of PROMs particularly given that the LSHTM work was commissioned by the Department of Health. In addition to central support, NHS commissioners have a pivotal role in promoting PROMs, particularly if they commission services contingent on collection of data.

Until recently, clinicians and providers have been wary that PROMs are being used to highlight the bad performers, rather than being used as a tool to improve the performance of particular clinical teams. A more positive emphasis is essential to gain clinician buy-in.

2. Measurement practicalities

PROMs can either be collected using generic instruments, which would enable comparison across different disease areas, or a disease-specific instrument which would only enable comparison within a therapeutic area. It very much depends on the purpose the PROMs are needed for. It would not be necessary for generic measure to be used to evaluate improvement in quality of care, particularly as disease-specific measures are more sensitive by their very nature. But generic measures may be the only option if the outcomes are to be incorporated into an overarching productivity measure.

The problems in choosing an instrument for collecting PROMs were demonstrated when BUPA initially used SF-36 as a generic instrument but then had to review its usage for cataract procedures due to the apparent insensitivity of the measure in that situation. BUPA now uses the specific VF-14 when assessing the success of cataract surgery. This can highlight the importance of choosing the correct tools and not just to stick to generic measures. On the other hand other experts feel that if the use of treatment/procedure has a low impact on the SF-36, may mean that the value of these procedures should be questioned. The more pragmatic see that these instruments are not designed to evaluate all procedures.

PROMs do not necessarily pick up issues such as adverse outcomes and complications from surgery. A secondary questionnaire concerning these issues may need to be circulated to circumvent this problem, and some information also has to be collected to risk adjust when comparing one hospital to another.

Concerning recruitment of patients, it is best undertaken on admission for surgical patients rather than at the pre-op assessment stage, for several reasons. At the pre-op stage it is not necessarily known for sure whether the operation will take place or where. Without assessment immediately prior to the procedure it would be impossible to understand the benefits of a particular procedure. Local training needs to be arranged for the staff who recruit patients. When the follow-up questionnaires are sent to the patients they should be sent by a third party, rather than the health institution concerned as patients will feel more able to provide honest and accurate answers. During all these steps of the measurement process, it is important to have the patients fully engaged.

Independent contractors such as CHKS suffer from a lack of central diktat to collect this information. Hospitals may be enthusiastic but do not want it impacting excessively on their staff's time. But if outside contractors were to take the responsibility away from the hospital, then the consequently reduced visibility of the collection process could be detrimental. Hence there is a difficult balancing act to be maintained.

Other measurement difficulties concern patients with chronic diseases. A longer-term perspective is required in these cases and not just measurement three months after an intervention. When CHKS explored monitoring chronic diseases they undertook measurement on an annual basis, but no consensus has been reached for the monitoring interval. To make matters even more complicated, many of those suffering from long-term conditions have multiple conditions and/or receive multiple treatments. This can present difficulties when deciphering which treatment has affected their quality of life to what extent, i.e. the association of observed health improvements to a particular intervention can be a complex link.

In other areas, such as mental health, there have already been other data collection exercises in place. HoNoS scores have been part of the Mental Health Minimum Dataset for a number of years, but despite that only a few providers collect HoNoS scores regularly, and they are collected by clinicians rather than patients and so are not strictly PROMs. The CORE-OM is another measure used in mental health which has been developed from the bottom up and can be recorded electronically, which may potentially be a useful PROM.

3. Experience of working with PROMs

One of the advantages of collecting PROMs is to be able to compare performance across institutions and even down to clinician level. BUPA has much experience in using PROMs within its hospitals and has decided hitherto to maintain the anonymity of clinicians. Although individual identification is becoming more acceptable, comparisons across centres between surgeons may not be helpful as it is the performance of the whole clinical team at centre level that may matter more.

At first there was a perception among some clinicians that the use of PROMs was a punitive measure (as discussed earlier) but clinicians have now gained more confidence in the process and attitudes have also changed following the problems at the Bristol Royal Infirmary. Comparative data are looked at centrally within BUPA, but there is an acknowledgement that the SF-36 data cannot be used alone when assessing performance. Hospital level benchmarking does occur within BUPA – you can look by condition and can compare outcomes. The key message is to keep the data as simple as possible.

Although clinicians are beginning to understand the role of PROMs, they continue to have concerns about how PROMs should be used alongside other important outcome measures such as biochemical test results, avoidance of mortality, symptom relief. All of these measures are important and reflect different perceptions of success.

Patient experience needs to be taken together with outcomes data to give a holistic view of the success of the health service. BUPA currently collects a lot of data on patient experience but it is analysed separately from the PROMs data. Patient expectations can have a profound effect. When expectations are more realistic, there is some evidence of better outcomes being recorded. Therefore it is important to use PROMs to describe the likely health gain from a treatment or procedure to inform patients and manage their expectations.

When using PROMs as performance management tools it should be noted they can only identify where there may be a problem and highlight areas that warrant further investigation. Thus PROMs raise questions but do not necessarily give the answers. However, collecting PROMs is an important first step in the process of systematic performance management.

4. Policy issues and where do we go from here

Listed below are a number of policy issues where PROMs may have a significant impact:

- o Patient's choice – patient-driven quality
- o Commissioning for outcomes
- o Evaluation of new technologies
- o Monitoring chronic conditions
- o Performance management
- o Consultant validation
- o Productivity measurement

Patient choice and patient-driven quality were previously seen as the main levers for the use of PROMs but as more emphasis is placed on commissioning for outcomes in the NHS and the monitoring of chronic diseases, they will also become important applications of PROMs. There is little evidence on the use of PROMs to stimulate choice. The most closely-related literature comes from the US and investigates the public disclosure of mortality data, but there the major impact was on clinicians and managers rather than patients and the public. It may only need a few patients to start analysing and understanding the data for PROMs to have a positive effect on patient care.

However there needs to be acknowledgement that there is an acceptable range for performance and attaining scores below average does not automatically imply bad performance.

If it were possible to link PROMs and other outcome measures to hospitalisation data, then measuring productivity and performance would be a much more feasible task. In addition, the monitoring of health status through collection of PROMs may be used to assess the impact of delayed treatment, i.e. would a delayed treatment mean that a patient could not achieve the expected outcomes. This data can also be used to influence the delivering and organising of services

Clinicians need to be interested in PROMs for collection of those data to work in practice. Clinicians need to see how PROMs data will be beneficial to the health service, to know how to act on the results of data collection, and to be able to incorporate such data collection into routine practice. One option to aid achievement in this regard would be to incorporate such activities into postgraduate training. Another option would be to use this data in physician revalidation appraisals, which could be used as an incentive to improve patient care.

Clinicians are not the only stakeholders who need to be incentivised. For PROMs measurement to work, managers must be involved in and actively promote the process. If PROMs collection were to be included in Healthcare Commission targets, then managers and commissioners would be more inclined to push for their collection. BUPA has a management bonus scheme in place to ensure adequate collection.

The cost of undertaking PROMs collection per patient is estimated by LSHTM at around £6, including 20% recruitment cost, and at around £3 per patient by BUPA, although increased automation will decrease the cost.

A lot of discussion in previous sections has concerned the collection of PROMs measures before and after a procedure. However PROMs, may also be useful if they are collected at referral, to inform clinical discussion in the outpatient setting. For example, primary care clinicians could include the PROMs value in the referral letter, thereby giving useful information to help the decision on whether to undertake a procedure or not. Allied to this, PROMs may also be a potential mechanism to subdivide those groups of patients to get an understanding of those most or least likely to benefit from an intervention.

If the data are collected they could be used to feed into the development of informed decision-making tools, alongside other biochemical markers. For instance, if adequate IT software were developed, all of these different types of data could be plugged into computer algorithms to help patients and their clinicians choose treatment pathways together.

Appendix 11 – Using Patient Outcomes Data in NHS Commissioning

Summary of OHE Commission Workshop

12th October 2007

This appendix summarises the discussion of a second workshop arranged to look at how the NHS could pragmatically and routinely collect patient outcome measures to assess the performance of the Health Service both at the local level and a more aggregated level.

In the first half of the workshop the research commissioned by the OHE Commission reviewing outcome measures used within mental health, chronic obstructive pulmonary disease (COPD) and colorectal cancer was presented by the respective authors. In the second half of the workshop four different stakeholders from different parts of the NHS and the Department of Health were asked to give their views on how these measures could be used within the NHS.

1. OHE Commission Research

1.1 Mental Health – Rowena Jacobs, University of York

Rowena Jacobs presented her work on Outcome Measures in Mental Health (a detailed summary of her work is presented in Appendix 2 and the full report on the OHE Commission website). The two main measures described were a clinician-rated measure - HoNOS, which is mandated by the Mental Health Minimum Data Set and a patient-rated measures – CORE-OM.

In mental health after a hiatus there is once again a policy push for outcome measurement with a compendium of outcome measures being put together. However, data collection is not widespread due to a number of barriers. Clinicians may not see the clinical benefits of the measures as the scores are not fed back systematically, which hinders compliance significantly. There is a need for an outcome-orientated clinician culture (discussed in great length in the earlier workshop – see Appendix 10).

A second barrier is the lack of adequate IT systems and IT staff and IT training within the clinical workforce.

This research also explored whether NHS commissioning could be an incentive for the collection of outcome measures. Although commissioners were positive about this driver, providers were less so. Similarly there was uncertainty over whether incorporating patient outcomes into Payment by Results (PbR) would deliver service improvement. A more viable lever may be to incorporate outcomes measurement into Healthcare Commission standards.

The audience endorsed the need for better information as a driver for service improvement. Information centres could be set up that publish data on quality by provider.

1.2 COPD – Andrew Briggs, University of Glasgow

Andrew Briggs presented the work he has undertaken with Helen Starkie at the University of Glasgow (a detailed summary of the research is presented in Appendix 3). There are number of outcomes measures used within COPD. These can be grouped into hard end-points (such as

mortality and emergency hospitalisations), clinical measures (such as forced expiratory volume in one second [FEV1]), and patients' health outcome measures.

A range of generic health outcome measures have been used to monitor COPD treatments and progression including the SF-36 and the EQ-5D. A number of disease-specific health outcome measures have also been used including the Chronic Respiratory Questionnaire and the St George's Respiratory Questionnaire.

Although clinical measures are routinely recorded, there seems to be less coverage of quality-of-life outcome measures particularly outside the clinical trial environment. It is unlikely that COPD patients will be fit enough to fill in a questionnaire on admission to hospital. There are also barriers to clinician administration as discussed under mental health, i.e. the need for clinicians to see the clinical benefits of using these measures and have the results immediately available. One solution may be to have more sophisticated IT systems which could involve touch-screen questionnaires in GP surgeries which would allow data to be transferred quickly into the patient record.

To assess the productivity of the health service over time, we need to have access to time series of data. A database in Cardiff is collecting EQ-5D for a number of different conditions where the data are collected six weeks after admission at a GP appointment. This may not be relevant timing for COPD as the exacerbation, where QALYs would have been lost, would already have occurred. An alternative proposal is to estimate the average number of QALYs lost per exacerbations and simply count the number of exacerbations. However, not all were in agreement that this was the most feasible way forward as they were wary about such a surrogate marker of utility.

1.3 Colorectal Cancer – Paul Trueman, York Health Economics Consortium

Paul Trueman presented his work on outcome measures in colorectal cancer (a detailed summary of his work is presented in Appendix 4 and the full report on the OHE Commission website). This work built upon an earlier project commissioned by the Department of Health, looking at costs and outcomes associated with colorectal cancer.

In identifying the different measures in use, a search of cancer registries was undertaken in addition to a literature review. This was supplemented by discussions with various stakeholders. As five-year survival rates are only around 50% for colorectal cancer patients, survival/mortality data are particularly meaningful measures of outcomes in this disease area, and may also be useful in the measurement of productivity. However there have been few attempts within the cancer registries to capture quality of life information.

From the literature review five quality-of-life outcome measures were identified. Two versions of the EORTC instruments, one a questionnaire for cancer generally and the second one specifically adapted for colorectal cancer. FACT-C is another instrument specific for colorectal cancer patients. The other two identified were generic health outcomes instruments already discussed in relation to COPD.

Although mortality data collection is essential to assess productivity improvements, the routine collection of patient-reported outcome measures would be a useful supplement. Paul Trueman recommended that the use of a generic instrument such as EQ-5D would be a good adjunct particularly during routine consultations. Indeed it would be easy to incorporate the EQ-5D into the patient experience questionnaire. If these measures are to be used primarily for the measurement of productivity it may be that sampling would reduce the burden further, but if the objective is to be patient monitoring tool then sampling is not necessary.

The second half of the workshop focused on the NHS viewpoint. Four speakers from different parts of the NHS were invited to give their thoughts on the research presented in the first half and how important outcomes measurement may be to measuring the performance and productivity of the NHS.

2. NHS PERSPECTIVE

2.1 Shane Gordon, Colchester Practice-Based Commissioning Group

A lot of decision-making by patients is facilitated by the interaction with the clinician. Patient-reported outcome measures may be an important input into high quality discussion and may enable patients to have realistic expectations of an intervention and the health service in general. These measures may well inform patient choice. This is however only one of many pieces of information that may be part of the decision-making process, and should be placed in context with other information available. For example mortality may be important when looking at cancer interventions however in palliative care quality of life is much more important than mortality.

Patient reported outcome measures may also be used for comparative purposes. Clinicians would be wary of this, particularly if there is no casemix adjustment, but these data could be very useful for commissioners when deciding which services to purchase, and in performance management. There is also potential to include outcomes measurement in the Quality and Outcomes Framework. Nationally consistent use of outcomes measures is required. This will permit comparison and benchmarking.

Other foreseeable problems include patients suffering from questionnaire fatigue, the administrative burden in collecting data, the difficulty in physically populating the dataset and the capacity of IT systems to deal with the quantity of data. As has been stated earlier, the outcomes data need to be analysed at several levels: individual level, clinician level, practice level, commissioner level, SHA level and national level. At many of these levels the skills and resources to analyse the data do not yet exist, hence there needs to be serious capacity planning for health outcome measurement to be used successfully to monitor productivity and performance.

2.2 Daphne Austin, Consultant in Public Health, West Midlands Specialised Services Agency

Both epidemiological data in terms of health status and productivity data in terms of what the NHS gets for its money have been discussed. For each of these purposes the data have to be paid for – good quality will cost a substantial amount of money – and even at regional level there is currently a lack of information specialists and epidemiologists to make sense of the data.

There are four general questions that NHS commissioners want to be able to answer:

- 1 What health care should be made available to patients and with what aim?
- 2 Do interventions fulfil their promise, i.e. what are we actually getting?
- 3 What are we getting from this provider?
- 4 What are we providing to this patient?

Practice-based commissioners are more likely to focus on questions 3 and 4, whilst planners at a high level may be more interested in 1 and 2. Different levels of data are required for these different perspectives. For example for treatments such as the ultra orphan drugs, investment decisions may need to be made at the individual patient level.

The Atkinson Review advocates moving from activity measures to patient-reported outcome measures, however proxy measures may also be helpful to a commissioner, particularly as they are easy to collect and therefore can be rapidly fed back. Proxy measures are valid outcomes of interest to society in terms of funding and productivity and therefore should not be disregarded so easily.

The measure used when evaluating an intervention has to be fit for the particular question, hence when a clinical trial is carried out the health gain of interest should influence the actual focus of the research. There are real problems in the outcome measures that are currently used and whether they can really tell us whether we are getting the appropriate result from an intervention.

Collection of outcomes data could be used to re-evaluate decisions to see whether we are getting value for money in the real world, e.g. post-implementation of NICE guidance. Therefore there may be a role for longitudinal data and the compiling of large databases to help inform investment decisions.

Outcome measures from a commissioner perspective should be about informing decision-making, priority setting and service specification. How should commissioners collectively make decisions about what should be provided by the NHS? For example, recombinant factor VIII which is purchased for the use in haemophiliacs is prescribed over a range of different doses depending on the clinician in charge. There does not seem to be a standardised dosage schedule to be titrated against. The variation in doses can make a major difference in the costs of the treatment – this could be an area where outcome measurement is prioritised.

Not everyone is enthusiastic about the QALY. It may be a good way of screening the benefits of an intervention, but it does not distinguish between 365 patients getting an extra day of life each and one patient getting an extra year of life. Although cost-effectiveness is an invaluable concept, the QALY needs to be followed up with detail when making commissioning decisions. Many commissioners would simply not make a funding decision on the basis of a QALY number. Cost-effectiveness ratios do not take account of the affordability dimension. Opportunity cost from a commissioner perspective is not just an abstract concept. In practice, investment in expensive but 'cost-effective' technologies means that other parts of population may miss out on valued treatments.

In conclusion, there needs to be a change in how investment decisions are made. There are a number of different perspectives that have to be accounted for in these decisions, e.g. commissioners, public health, and general management. Alongside this there needs to be more active engagement of patients in the process, and capacity and capability in this aspect needs to be built up.

2.3 Joe Rafferty, Director of Commissioning, North West Strategic Health Authority

This is a North West perspective on patient outcome measurement. There does indeed need to be a move towards the measurement of outcomes rather than processes if the aim is for reduced inequality, improved quality and safety, improved value for money. Until now there has been focus on efficiency at Strategic Health Authority level. It is now important to put quality and outcomes at the heart of the agenda, whilst recognising that quality without efficiency is unsustainable.

In North West England, NHS commissioners are in charge of £12 billion of public money and the return on this investment has to be more effectively and explicitly measured. A number of exercises have been undertaken to work towards the measurement of quality and outcomes, which have been linked to the implementation nationally of 'payment by results'.

Quality needs to be measured in a consistent way, and there needs to be rigorous standardisation of data on quality before it can be reported to the public. Incentives are the cornerstone of behaviour and therefore in the North West an initiative around payment for quality has been introduced, based on the Premier 'pay for performance' initiative in the US. For example, a clinical pathway was set out for Coronary Artery Bypass Grafts (CABGs). If 50% or less of these clinical steps are adhered to then mortality increases significantly, hence a reward has been put in place to encourage adherence in the US. Incentives do not always have to be financial – some PCTs have stated that they would rather have recognition as a reward for excelling.

Premier used incentives in five high-volume conditions: acute myocardial infarction, CABG, heart failure, community acquired pneumonia, hip and knee replacement. In the North West this work is being replicated. There is recognition that an initiative such as this needs to be co-ordinated at SHA level at least at the beginning, and there are established agreements and local governance arrangements with each of the PCTs. PCTs had to put some of their growth monies into this and therefore it was important to obtain their buy-in.

In order to have the right skills mix to move this initiative forward North West SHA are in the process of finding the right partner. At present clinical outcomes will be used as the primary driver for improvement but including work on patient reported outcome measures will also be important, particularly as this would assess the well-being of the patient. It may be necessary however to put greater emphasis on clinical outcome to ensure clinician buy-in. The key objective is to build a structured systematic picture by benchmarking for clinical and other patient outcomes on an industrial scale so that we are able to commission for outcomes.

Clinicians are now beginning to see the importance of such initiatives and want to drive this forward. There has been an astonishing level of clinical sign-up – they are interested in quality incentives. For example the cardiac surgeons see it as an opportunity to benchmark themselves against their North American counterparts. This project is voluntary and not mandatory, but the adoption rate is high with over 60 providers and commissioners involved. Many see benefits in participation, i.e. clear and consistent objectives for creating value.

To demonstrate the scope for using such an exercise to improve quality, it is useful to look at the results of a recent audit carried out by the Lancashire and Cumbria Cardiac Network. They undertook an audit of patients to evaluate how many patients were receiving seven evidence based interventions as listed below:

1. Early aspirin
2. Early beta blocker
3. Aspirin at discharge
4. Beta blocker at discharge
5. ACE inhibitor at discharge
6. Timely reperfusion
7. Smoking cessation

Only 32% of patients received all seven, which shows the scope for performance improvement even in a therapeutic area where there has been significant investment.

The payment for quality scheme should commence as early as October 2008 with the first year of payment in October 2009. Although the project focuses on only five treatment areas to begin with, the intention is to link the outcomes information to needs assessment so that a real strategic view can be formed on where to invest. This can be extended to other therapeutic areas but there needs to be detailed work undertaken on where opportunities may lie, i.e. where appropriate data already exists or are easy to collect. Mental health may therefore not be an appropriate area at present.

A system to gauge health care gain needs to be put in place, and if there is a transparent enough measure of quality alongside, the patient choice agenda may also be driven by it. The commissioning system needs to evolve so that it is more consistent with public value and with demonstrable outcome benefits. The initiative being put in place in the North West may be one step in this direction.

2.4 Bob Ricketts, Head of Demand Side Reform, Department of Health

Free choice for patients of provider of elective surgery comes into place in April 2008. As patients decide not to use services that are seen as underperforming, there will be changes in referral patterns. Alongside this change, the Department of Health are trying to provide quality improvement levers to PCTs. Patient reported outcome measures are an important part of this.

PROMs may be useful in terms of setting priorities and evaluating return on investment, e.g. are we getting the appropriate improvement in quality of life from elective surgery procedures? Commissioners could also use these measures to lever quality from contracts, and if they were presented in a usable way patients could use the information for decisions around choice. But is there really evidence that PROMs can be used in all these ways?

There may well be viable alternatives to PROMs but PROMs have been the measures used most consistently. The Department of Health needs to understand the utility, feasibility, affordability and the implementation issues of PROMs before introducing them into the NHS operating framework on a large scale. Thus far PROMs have not been used for setting priorities in the UK. There is limited evidence from research presented earlier that PROMs can lever improvements, but there are also issues around the skillset of commissioners to be able to analyse such data.

Some commissioners are already building outcomes into contracts with an aim of levering improvements. The PROMs have proven methodologies. Nick Black and his team at the LSHTM have undertaken pilots on behalf of the Department of Health on the use of PROMS for elective surgery. The pilots show that such an exercise could be successful. In a parallel study undertaken by Ray Fitzpatrick (University of Oxford) and colleagues in six long-term conditions, it has been shown that such data collection could provide rich information sources even within more complex fields. If the data are of sufficient quality, particularly at individual level, they could be used routinely in discussions with patients.

The focus should not just be on the commissioner, though. Providers are just as likely to drive the data collection exercise, particularly if they have a good clinical team and want to prove their value. But there needs to be adequate investment in presentation of data so that it is useful to all interested audiences.

If a mandatory PROMs scheme were to be implemented through the national contract for elective surgery procedures as recommended by Nick Black (see Appendix 1), it may cost about £10 per procedure, which amounts to about £700k overall. There needs to be discussion over whose responsibility it is to pay for implementation – should providers absorb the cost to prove they are serious about monitoring quality or should the Department of Health pay for it? An even bigger issue is who undertakes the analysis. It may be efficient for the Department of Health to do this centrally, but to ensure that this is taken forward, national contracts may be the most appropriate avenue to incentivise analysis by commissioners.

It may be best that PROMs collection be put in National Contracts on a voluntary basis at first, but by 2009 or 2010 it may be feasible to make it mandatory. In the voluntary years, all organisations will be able to learn and there will be opportunities for leaders to demonstrate what they are doing,

as long as the direction of travel in the NHS is explicit and a serious commitment is made. In addition, data should be made public for high volume elective surgical procedures. This should be extended into other areas such as long-term conditions where there may well be greater gains for patients and commissioners.

The practicalities of collecting PROMs data need a lot of thought. It may be useful to learn from the experience of the patient experience survey, where 1.5 million patients have been involved. Organisational level surveys with outputs at trust level may not help patients or clinicians as a Trust may not assess a patient before and after a procedure. The focus should be on one or two conditions at first.

2.5 Summary and discussion led by David Stout, NHS Confederation PCT Network

There was a heartening consistency throughout the discussions. The NHS should care about measuring quality, but not enough work has been done on it thus far. In the first half of this workshop, the focus was mainly on the types of measures that could be used to assess quality and the evidence base for them, ranging from hard clinical endpoints to patient reported outcome measures and patient experience surveys. There are some robust data already collected through clinical trials and robust instruments to measure quality across the therapeutic areas discussed.

The discussion presented the technical difficulties, but the consistent view from the first half of the workshop is that we may be able to use generic measures of self-reported outcomes. However there needs to be consistency in usage nationally

The outcome measures have to be practical and meaningful to the commissioner, the clinician and the patient. They need to be collected for the right purpose – targeted at the questions that need answering rather than those that sound theoretically interesting. The data could be used to better inform patients, for the quality assurance of providers, and helping to make priority decisions on investment – but all stakeholders need to be clear about the purpose.

There will need to be investment not only in data collection and analysis but also in feedback. For this worthwhile exercise to be truly taken forward, resources need to be earmarked (for IT infrastructure and appropriate analytical skills), particularly if consistency nationally is to be achieved.

Contact Us:

Office of Health Economics
12 Whitehall
London
SW1A 2DY
UK

Website: www.ohe.org
Tel: +44 (0)20 7747 8850
Fax: +44 (0)20 7747 8851
E-mail: ohgeneral@ohe.org