

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
and funded by Dr Foster Limited

Written by

John Appleby
Nancy Devlin

Editorial board

Alan Maynard
Helen Scott
Andrew Vallance-Owen

November 2004

Contents

Preface

Executive summary

- 1 Putting the “health” back into healthcare management**
- 2 What can be gained by measuring improvements in health?**
Improving the performance of providers
 - Directly managing the performance of doctors: revalidation and professional regulation
 - Directly managing the performance of hospitals: targets
 - Indirect performance management: patient choice
 - Indirect performance management: Payment by results – what results?
 System-wide opportunities for quality improvement
 - Health technology (re)evaluation in the real world
 - Patients’ values
 - Clinical thresholds
 - Fairness and healthcare needs at the population level
 - Measuring productivity
 Conclusions
- 3 How should we measure changes in health?**
Condition-specific measures
Generic measures
Existing routine use of HRQoL in the UK
When should HRQoL be measured?
How should the data be reported?
Conclusions
- 4 What are the risks and costs?**
How much will it cost?
 - Language
 - Method of administration
 - Sample size and coverage requirements
 4.1 Will variation in performance be detected?
4.2 Clinician support
4.3 Gaming
4.4 Will the information be used? Will it result in improvement?
4.5 Conclusions
- 5 What next?**
Analysing existing data
Recommendations for a pilot study
Co-ordination of HRQoL initiative across the NHS
- 6 Conclusions: Time for action**

Appendices

Preface

The Dr Foster Ethics committee is a regulatory body, totally independent of Dr Foster, whose role is to monitor the work of Dr Foster and ensure that the company presents data accurately and in an unbiased fashion. The committee investigates any complaints made to it about Dr Foster and endeavours to ensure that they are resolved to the satisfaction of all parties.

More recently the committee has been concerned with areas it deems to be of importance to the delivery of high quality healthcare, patient choice and consumer protection. This paper, commissioned from Prof John Appleby and Prof Nancy Devlin, is the first project of this programme. It clearly demonstrates that the systems for measuring success in terms of improving patient quality of health are possible and merit further evaluation and use in the health service.

This report was made possible by a commission from the Dr Foster Ethics Committee in May 2004. In undertaking this work, the authors consulted with the following people, whose assistance is noted with much thanks.

- Professor Sir Ara Darzi, Imperial College
- The EuroQoL Group
- Dr Alastair Fischer, Alec Miners & Louise Longworth, NICE
- Professor Bill Gillespie, Hull and York Medical School
- Dr Simon Dixon, Sheffield Health Economics Group
- Dr John Fox, Head of Statistics Division, Department of Health
- Simon Stevens, Prime Minister's adviser on health
- The Royal Colleges were invited to comment: detailed responses were received

from:

Royal College of Psychiatrists

Royal College of Ophthalmologists

Royal College of Orthopaedics

Royal College of Midwives

Royal College of Paediatrics and Child Health

- Members of the Dr Foster Ethics Committee:

Dr Jack Tinker (chair), emeritus dean, Royal Society of Medicine

Sir Donald Irvine (vice chair), former president, General Medical Council

Dr Michael Dixon, chair, NHS Alliance

Dr Andrew Vallance-Owen, Group Medical Director, BUPA

Dianne Hayter, vice chairman, Financial Services Consumer Panel; board member, National Consumer Council, board member (until early 2004), National Patient Safety Agency; former chief executive of two health charities (Alcohol Concern and Pelican Cancer Centre); member of the Labour Party NEC

Professor Alan Maynard, director, Health Policy Unit, York University; chair, York Health Services NHS Trust

Bridget Gill, head of communication, North and East Yorkshire & Northern Lincolnshire SHA (IHM representative)

Trevor Campbell Davis, chief executive, Oxford Radcliffe NHS Trust

Douglas Webb, chief executive, Cornwall Care

Dr Philip Davies, medical director, Pontypridd and Rhondda NHS Trust

Helen Scott, editor, British Journal of Nursing

Linda Buckingham, director, Unit of Health Services Management,
Hull University Business School

Dr Helen Earl, consultant oncologist, Addenbrooke's NHS Trust

Professor Christopher Moran, professor of trauma and orthopaedic surgery,
Queen's Medical Centre, University Hospital, Nottingham

Professor Barry McCormick, director of economics and operational research,
Department of Health (Observer)

Remaining oversights or errors are, of course, the authors' responsibility – and opinions expressed in this paper are the authors' own.

John Appleby is Chief Economist at the King's Fund and Visiting Professor at City University. **Nancy Devlin** is Professor of Health Economics at City University and is a visiting fellow at the King's Fund.

The King's Fund is an independent charitable foundation working for better health, especially in London. The Fund carries out research, policy analysis and development activities, working on its own, in partnerships, and through grants. The King's Fund offers a major resource to people working in health, providing leadership and education courses; seminars and workshops; publications; information and library services; and conference and meeting facilities.

City University is "The University for business and the professions". The City Health Economics Centre at City University is a small London-based research group undertaking innovative research of direct policy relevance.

For more information about this document, please contact:

John Appleby

Chief Economist

King's Fund

11–13 Cavendish Square

London W1G 0AN

jappleby@kingsfund.org.uk

Executive summary

In five years' time total spending on all healthcare services in the UK is likely to reach a staggering £144 billion – with NHS spending rising to £128 billion – a threefold increase since 1997. In other words, by 2009, just over £1 in every 10 circulating in the entire economy will be devoted to healthcare.

What we get for our healthcare spending has always been an important question but, as more and more of the nation's wealth is devoted to healthcare, answering this question becomes increasingly urgent.

Given the immense effort, resources and, indeed, anguish expended over recent decades on reducing waiting times and increasing activity, the public and patients may be forgiven for thinking that the main performance goal of the NHS is to provide quick treatment. But the primary goal of the health service is to improve people's health, and a fundamental goal within this is to improve patients' quality of life.

The answer to “what do we get for our healthcare spending?” should be measured in terms of health, but the NHS – like all health systems around the world – has very little idea of what it produces as it does not routinely measure patients' Health Related Quality of Life (HRQoL).

It is becoming increasingly obvious that this failure stands in the way of establishing where the real gains in health are to be made, which treatments and modes of delivery are most effective, how hospitals, clinical teams and, indeed, clinicians are really performing, and where the extra money we plan to spend on the NHS would produce the best results.

The potential benefits of routinely measuring patient-assessed HRQoL are wide ranging, providing basic evidence to inform the revalidation of clinicians, the performance management of hospitals, and real patient choice. Such data would have wider system impacts, helping in the re-evaluation of old and new treatments as they are delivered in the real world, providing a source of patients' own views about health and healthcare, feeding into current debates and research into the best ways to measure NHS productivity, and the tracking of changes in clinical opinion and action regarding when to admit and treat patients.

Linking information on patient-assessed HRQoL to patients' medical records and other data sets would, for the first time, allow proper evaluation of broader government health policies as they affect equity: not just how much benefit, but who benefits.

The failure to measure an absolutely fundamental outcome of healthcare has not been due to the lack of ways of measuring health related quality of life; measures have existed for many years, have been intensely researched, and have provided well recognised and accepted measures of outcome in clinical trials for decades. Thousands of disease- and patient group-specific self-completed questionnaires have been developed and tested along with generic measures such as the SF-36[®] and the EQ-5D.

This paper concludes that there is much to be gained by routine before-and-after measurements of health, both as a means of monitoring and managing the performance of providers, and as a means of facilitating a system-wide refocusing of the NHS on health.

Overall, having reviewed evidence of the potential benefits and costs of routine measurement of outcomes, current knowledge and experience of how to produce this information, and current clinical and technical opinion on this matter, more information is required before the routine use of HRQoL can be advocated. The bottom line in considering routine use of HRQoL to manage the performance of NHS providers is not only whether this will make a positive difference to performance and the well-being of patients, but also the scale of this difference. In short, will the benefits outweigh the costs involved?

Three issues need further investigation:

- Which HRQoL measures best discriminate between good and poor providers?
- How often, when, and how much would it cost to collect and analyse this data?
- How can the information be presented so that it generates action – and what improvements will result?

1

Putting the “health” back into healthcare management

More than 100 years ago, Florence Nightingale devised a simple three-point health-related outcome measure for her patients – relieved, unrelieved and dead (Nightingale 1863). Since then, the NHS (indeed, all healthcare systems) has failed even to match this relatively low level of sophistication in measuring the impact of its services on patients – and, by extension, the performance of the health system as a whole. Indeed, for many years the NHS reported its hospital activity statistics as a composite of two measures – discharges *and* deaths – making no overt distinction between the two. Today, along with access and process measures of performance, such as waiting times and length of stay, the main output measures for the NHS are finished consultant episodes and the (somewhat Harry Potteresque) patient “spell”. Such statistical disregard for what can only be considered an important (if not *the most* important) measure of performance for the NHS – patients’ *health* – may seem callous. Of course, this is more apparent than real but, if so, why does the NHS continue to ignore such a crucial indicator of its performance?

Perhaps more puzzling is the fact that, since Florence Nightingale, clinicians, trialists, psychologists, economists and many others have been involved in developing literally thousands of disease-specific, patient group-specific and generic tools to measure health status as essential measures of the health outcomes in clinical evaluations. The past 30 years in particular have seen a tremendous growth in the development and use of patient-assessed health-related quality of life (HRQoL) measures. This reflects in part a greater recognition that *biomedical* measures of health status (for example, blood pressure, tumour size, mortality) fail to fully capture what individuals and patients feel is important, which, in turn, has reflected moves towards a more patient-centred approach in medical care and hence a shift to a more “*biopsychosocial*” perspective.

The NHS has been a direct beneficiary or “end-user” of this work (and the application of HRQoL measures) through, for example, the health technology assessment work of the National Institute of Clinical Excellence (NICE) and in general through the generation of evidence to inform best clinical practice. However, the NHS has been slow to explore the potential benefits of patient-assessed HRQoL in relation to, for example, policy and practice on performance management, resource allocation and budget setting – despite recognition that routine measurement of quality of life could provide useful information. For example, in 1991 the then Chief Economist at the Department of Health, Clive Smeed, noted a number of areas where systematic information on HRQoL would improve policy and decision-making, from setting the national NHS budget and priorities across all healthcare interventions (including, importantly, those which affected quality rather than length of life) to providing the basis for effective purchasing of care in the NHS and strategic and operational decisions within providers.

Nevertheless, he was less sanguine about the importance and impact information on quality of life would (or should) have on policy making, stating that such data would never be more than one factor to be taken into account when comparing policy options. In his view, quality of life information would always be decision *aiding*, not decision *making* (Smee 1991).

This highlights a view, based partly on the practicalities of (and disagreements over) the use of HRQoL in such combined outcome measures as Quality Adjusted Life Years (QALYs), which perhaps carried weight in 1991. Since then not only has the use and development of quality of life measures moved on, but so too has the policy environment, with new (patient choice) and old (productivity measurement in the NHS) issues emerging to which HRQoL information could make a significant contribution.

A more fundamental question for the NHS, however, is how it can ensure the delivery of one of its key goals – improving patients' quality of life (cf Secretary of State for Health 2000) – when it has no routine system for measuring the very thing it is trying to improve. As Kind and Williams (2004) argue: “The time has come for those with a responsibility for health policy to demonstrate that what really matters is patient outcomes. The failure to grasp this nettle has... [stood] in the way of a really radical rethink about how best to measure success in healthcare.”

As a contribution to moving the debate on, this paper reviews the benefits, costs and practical and technical issues associated with the routine measurement and use of patient-assessed HRQoL to manage the performance of healthcare providers.

Our focus is on the potential use of HRQoL to identify variations in performance between doctors, surgical units and hospitals, and the way in which such information might be used to lift poor performance – from current moves to the use of quasi-market mechanisms and patient choice to formal regulation/inspection and target setting. Throwing the focus of the NHS on to health improvement, and a commitment to measuring this, also has potential benefits beyond the management of provider performance. We consider the system-wide implications of using routinely collected HRQoL information, from monitoring and improving the care of individual patients to decisions about allocating resources.

First we outline the potential benefits of using HRQoL in the health sector. Then we describe the measurement tools available for this purpose, while also reviewing the current use of these instruments in the UK. Next we assess the risks and challenges of incorporating HRQoL data into the management of providers. We conclude with recommendations for research required to inform the routine collection and use of patient-assessed HRQoL in the NHS.

2

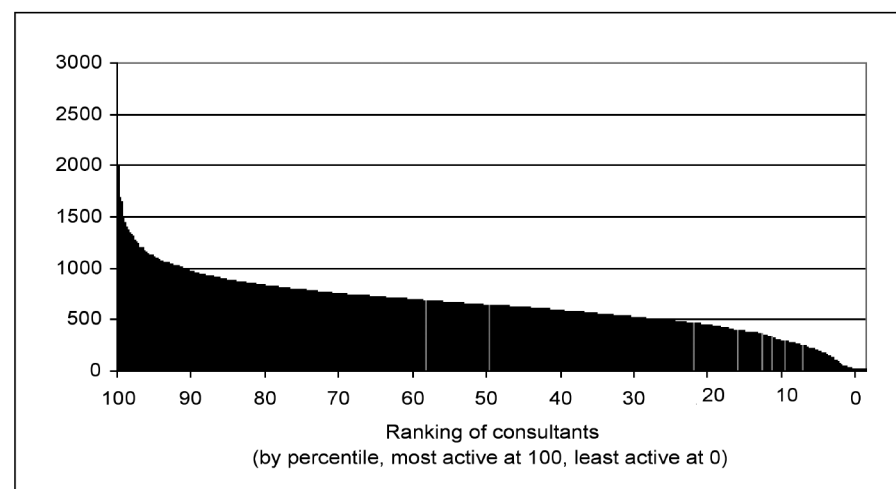
What can be gained by measuring improvements in health?

It may seem self evident that healthcare services and, more importantly, patients, have much to gain from the NHS re-focusing thinking about quality and performance firmly on to *health gain – as viewed by patients themselves*. However, it is worth setting out the potential benefits of this view in two ways. First in Improving the performance of providers below, we consider the use of HRQoL data in managing, either directly or indirectly, the performance of providers. The collection of HRQoL data for this purpose would generate a vast set of information on the health of patients before treatment and their health gains from treatment. Second, this data would provide powerful new insights into wider issues such as priority setting, resource allocation and strategic planning in the health sector. These benefits are described in System-wide opportunities for quality improvement, p11.

Improving the performance of providers

Two observations suggest that the performance of NHS providers is a cause for concern. First, there are wide variations in the performance of individual clinicians within the NHS. Figure 1 below suggests from existing resources the potential for substantial improvements in performance, but these observations are based on finished consultant episodes (FCEs) and, while this degree of variation in throughput is worrying, activity is not all that matters.

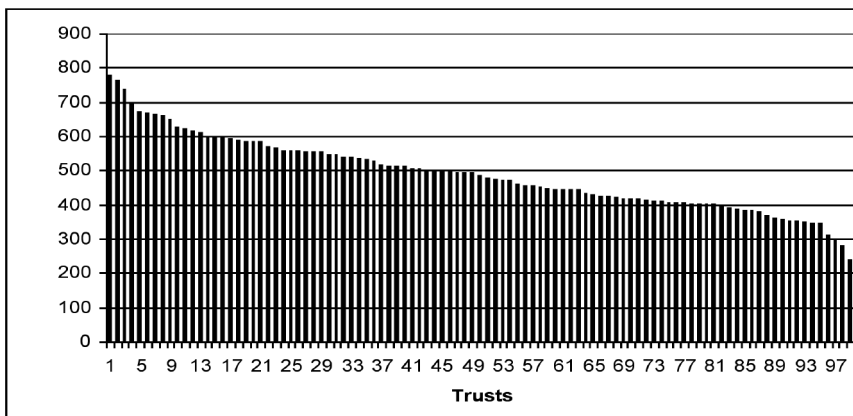
Figure 1. Ranked activity per consultant surgeon (FCEs); Trauma and Orthopaedics, 1999–2000



Source: Maynard and Bloor (2003)

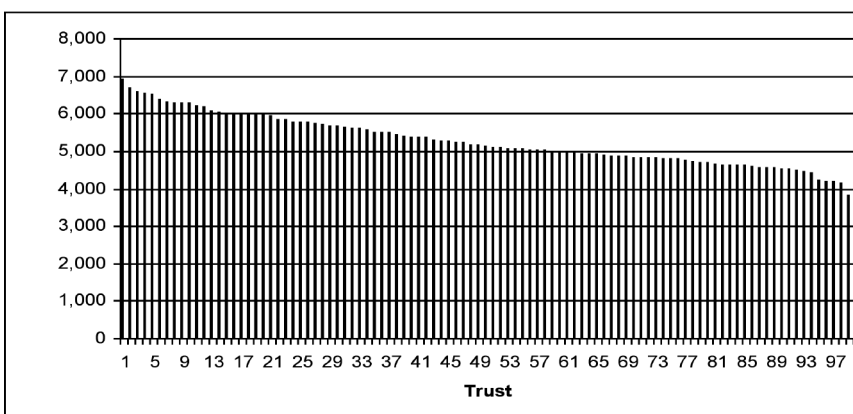
Second, there are wide and persistent variations in the performance of hospitals. But this is on crude “negative” measures of (adverse) outcomes, such as post-operative mortality and readmission rates (see, for example, figures 2 and 3) and performance on “intermediate” indicators of performance, such as waiting times (see, for example, figure 4).

Figure 2. Deaths in hospital following surgery: English trusts, 1998/99



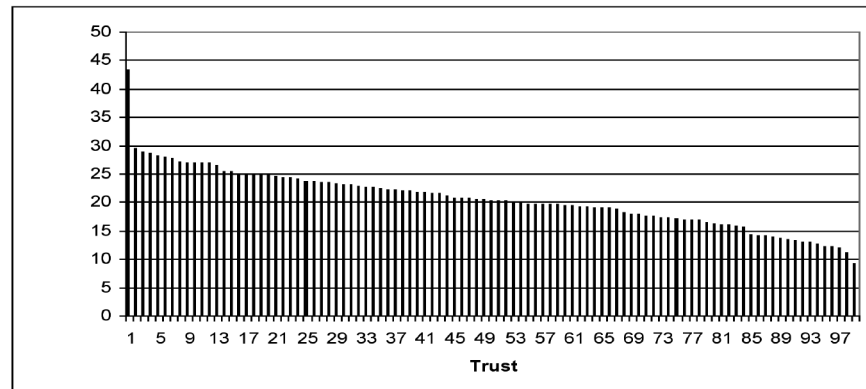
Source: Department of Health (2000)

Figure 3. Emergency readmission to hospital rates within 30 days of discharge: English trusts, 1998/1999



Source: Department of Health (2000)

Figure 4. Inpatient waiting list per 1,000 head of population: English trusts, 1998/99



Source: Department of Health (2000)

“Good quality healthcare” involves more than getting quick treatment that doesn’t kill you. *Prima facie*, there is a strong case for putting health improvement at the heart of performance management of all healthcare providers. There is a wide literature on measuring, managing, incentivising and accrediting the performance of healthcare providers – yet almost *none* of it incorporates health improvement. There is an even wider literature on the measurement of health states and health-related quality of life – yet in almost all cases the focus is on the effectiveness of *treatments*, not *those delivering them*.¹

Information on HRQoL could be used to manage the performance of providers in two ways (see Table 1 overleaf). We have used the term *direct* approaches to refer to planned actions that could be taken in response to evidence on performance – such as management or regulators directly rewarding or penalising good and poor performance respectively through pre-specified processes. *Indirect* approaches rely on a behavioural response to pressures that may emerge from evidence on performance – for example, if healthcare purchasers use information on performance to choose good performers over poor performers (*“caveat emptor”*).

1

A survey of the EuroQoL Group, May 2004, by the authors of this paper revealed that despite the vast array of studies using the EQ-5D to identify health gains following various treatments, very few of these identified individual healthcare providers in their data sets, and none (as far as we could determine) analysed differences between healthcare providers.

Table 1. Managing performance using measures of HRQoL

	Direct	Indirect
Good performance	<p>Evidence required to retain licence to practise through five-year revalidation and annual appraisal.</p> <p>Financial rewards (for example, bonuses) for exceptional performance.</p> <p>Evidence to inform consultants' Clinical Excellence Awards.</p> <p>Evidence to inform compliance with clinical contracts (for example, through negotiated "job plan").</p>	<p>Doctors strengthen clinical governance through the evidence they require for revalidation.</p> <p>Purchasers use the information to choose good quality providers: good performers gain market share and revenue.</p> <p>Patients, given a choice of provider at the point of referral, access and use the information to choose good quality providers: good performers gain market share and revenue.</p> <p>Enhanced professional reputation/satisfaction.</p>
Poor performance	<p>Annual appraisal and revalidation may show that remedial actions may be required, resulting in improved performance – or the licence to practice withdrawn.</p> <p>Disciplinary actions outside the revalidation process.</p>	<p>Loss of professional reputation/satisfaction.</p> <p>Purchasers and individual consumers use HRQoL information to avoid low quality: poor performers lose market share and revenue.</p>

Information on doctors'/hospitals' ability to improve their patients' HRQoL can link both to patients' and PCTs' choices (if buyers "vote with their feet"), which, in turn, can link to financial rewards for good performance. These links already exist: promoting patient choice and Payment by Results are key policy initiatives in the NHS. Further, processes for ensuring all doctors are fit to practice are already being put into place. What is lacking, in both instances, is a focus on *health*, as viewed by those receiving care.

Directly managing the performance of clinicians: revalidation and professional regulation

New systems for professional registration in medicine are currently being introduced. From 2005, doctors who wish to practise medicine in the UK will be required to hold a licence to practise – which is retained by "revalidating" at regular (five-yearly) intervals against the GMC's professional standards set out in the booklet "Good Medical Practice". Doctors must follow or demonstrate that they are "up-to-date and fit to practice" (General Medical Council 2004). For the vast majority of doctors who are employed, revalidation will rely on clinical governance processes already in place, including the introduction in 1999 of annual appraisal for doctors. To meet these requirements, and to demonstrate compliance with Good Medical Practice, all doctors are required routinely to collect and report data and information on their clinical performance. Licences are to be granted to all doctors by the end of 2004; revalidation will commence in Spring 2005, with those to be considered initially selected at random. Evidence to be submitted is to cover the period from April 2003 to the date revalidation takes place.

The principles of good medical practice and standards of "competence, care and conduct" are outlined by the General Medical Council (2001). Many of these principles and standards, and therefore, presumably, the evidence to be submitted via annual appraisal, focus on personal knowledge and skills and the quality in the *process* or *delivery* of care (rather than *health outcomes* from treatment). Evidence must also be submitted to verify that there are no concerns about the doctor's practice or performance. This might include, for example, information on adverse outcomes from treatment.

Evidence on the changes to patients' HRQoL following treatment could considerably strengthen these processes by shifting the focus to the doctor's achievement of positive health gain in patients – not simply the avoidance of adverse events.

This would confer important advantages: it would highlight and reward instances of very good performance; it would incorporate patients' views about their own health into the process; and, perhaps most importantly, information on patients' changes in HRQoL would allow the identification of poor performance that may simply be missed by focusing on (often fairly rare) adverse events such as readmission or mortality rates. Similar arguments can be made for the inclusion of HRQoL information into clinicians' job plans under the new consultant contract.

Directly managing the performance of hospitals: targets

The principal means by which the performance of NHS hospital trusts is directly managed has been through a system of targets, star ratings and individual reports on hospitals' clinical quality of care by the Commission for Healthcare Audit and Inspection (now the Healthcare Commission) reports. However, as the Prime Minister's former health adviser Simon Stevens has pointed out, this performance system is being supplemented (and in part replaced) by a more laissez-faire process involving light touch regulation (from bodies such as the Office of the Foundation Trust Regulator and the Healthcare Commission) and more market-oriented performance-improvement mechanisms such as patient choice of provider (Stevens 2004). Nevertheless, the setting of targets remains (inevitably) very important. The policy question has not been, therefore, whether to retain or abandon targets, but how to simplify and make more relevant what has over the past few years come to be seen as a rather cumbersome, disjointed and distorting process.

A start on this revamping of targets has already begun. In order to tackle the question of quality changes and to improve the measurement of cost efficiency, for example, the Department of Health and the Treasury have changed the nature of the NHS value-for-money target.

Traditionally, the Treasury has set this target, as part of its Public Service Agreement (PSA) measures (Her Majesty's Treasury 2002a), in terms of a percentage improvement in the cost-weighted efficiency index.

However, as set out in the 2002 PSA targets, the requirement now is for the NHS to achieve a one per cent increase in (a redefined measure of) *cost efficiency* and a one per cent increase in the *value of quality* (Her Majesty's Treasury 2002b). The new methodology for measuring the cost and quality elements of the new value for money target are set out in Box 1 opposite. While the recognition that extra NHS spending can contribute to better value for money through changes in the quality of the services it provides is a step forward, exactly how the NHS is to measure (and value in monetary terms) increases in quality remains somewhat opaque. The availability of HRQoL evidence could provide a way forward.

Box 1. A new way to measure cost and quality changes in the NHS

As part of the 2002 Public Service Agreement targets, the value-for-money target was changed to reflect spending to improve quality (rather than just the volume) of services and a new measure of cost efficiency was introduced (superseding the cost-weighted efficiency index). The new targets apply only to the hospital and community health services, exclude primary care services for the time being, and will apply for the first time in the 2003/4 financial year.

Cost efficiency

Cost efficiency will be measured in terms of changes in the *unit costs* of procedures (health care resource groups – HRGs) adjusted for the mix of cases and health service inflation. Increases in costs associated with increases in quality are subtracted from these, leaving an estimate of the change in cost efficiency.

Quality

Changes in quality will be primarily measured on the basis of the *value of lives* saved from reducing mortality following healthcare interventions. This measure will be developed over time to capture aspects of quality changes other than lives saved.

Indirect performance management: patient choice

While there are moves to specify smarter targets for the NHS, the parallel moves to introduce greater patient choice beg the question of how patients are expected to exercise their rights to choose, in particular the information they might expect to consult in order to make the “right” choice. Since the publication of the NHS Plan (Secretary of State for Health 2000), policy on patient choice has expanded rapidly. Improving the choices available to patients in the NHS is seen as a good thing in itself – and the extent to which patients actively take up the opportunity to choose is also compatible with key NHS objectives, including reductions in waiting times and improvements in quality. Following consultations on choice (for example, Department of Health 2003) and a number of pilot schemes around the country (for example, London Modernisation Board 2004), it is the intention to offer by December 2005 up to five choices to patients at the point of (GP) referral. While the choice on offer in the pilots has been essentially one of quicker treatment (as part of a multi-pronged strategy to reduce waiting times), if recent large reductions in elective inpatient waiting times resulting from this combined approach are sustained, then the “attributes” of their care that patients will increasingly value are likely to switch away from waiting times towards other factors – such as the quality of care.

The role of HRQoL in patient choice is based on the assumption that (1) consumers who engage in informed choice obtain better quality care and (2), if sufficient numbers exercise informed choice, providers will be motivated to improve – particularly where providers are motivated by the effects of consumers’ choices on market share and revenue (Hibbard 2003). Patient choice in the NHS is arguably motivated more by reducing waiting times than anything else – and its potential role in giving leverage to quality improvements is debatable. Hibbard (2003), commenting on consumer choice in the US, notes that:

Currently there is no evidence to support these assumptions about consumer behaviour. To date, there is only limited use of comparative information by consumers. Perhaps because little evidence exists that consumers are using quality in making choices, the link between consumer choice and quality improvement has also not been observed. (p62)

This does not suggest that it is impossible to use consumer choice to leverage quality but, as Hibbard goes on to suggest, that we need to find effective ways of supporting informed consumer choices.

Improving the *effective* choices available for patients requires readily available information on the nature of the options presented to patients and, crucially, on those aspects of their care that they value most (Appleby *et al* 2003). For example, the reputation of healthcare providers has been shown to be crucial to the uptake of choice in the London Patient Choice Project (Burge *et al* 2004). However, while information currently available to patients, such as individual performance indicators published by the Healthcare Commission (2004a), or composite measures such as star ratings (CHI 2002), or hospital reviews by the Healthcare Commission (2004b) or Dr Foster Hospital Guides (2004) provide some of the information patients may want to consider, much of this information is simply inappropriate for the decision at hand (for example, mortality rates – when most care is about quality, not length of life changes) or is presented at the wrong level (for example, hospital, not clinical team).

Indirect performance management: Payment by Results – what results?

Effective commissioning by primary care trusts (PCTs) relies upon their being able to differentiate between the performance of providers, and to select and contract with the highest quality providers at the prevailing, fixed (HRG) prices. “Quality”, in this context, seems currently to be dominated by, indeed almost exclusively focused on, waiting times. The purchasing focus remains throughput and the production of services, rather than their demonstrated impact on improving health. The new provider reimbursement system being rolled out under Payment by Results requires a greater focus on *results* (that is, outcomes-based purchasing) if the hoped for gains in quality

throughout the NHS are to be generated. As with patient choice, purchaser choice will increasingly need to be informed by information appropriate to purchasers' objectives to improve the quality of care. And again, as with patient choice, information on patient-assessed HRQoL clearly has a key part to play in differentiating the performance of providers and hence provides an important input to PCT purchasing decisions.

System-wide opportunities for quality improvement

The direct and indirect performance systems noted above are, of course, part of a larger set of processes the NHS employs to improve performance. Patient-assessed HRQoL can not only provide essential data to inform managerial, patient choice and PCT purchasing decisions, and the process of revalidation and ensuring value for money, but also, more broadly, inform system-wide performance improvement and other NHS objectives such as fairness in access and, arguably more importantly, fairness in outcome of care. Below we note five system-wide opportunities for quality improvement: the first two are linked to the technology assessment work of the National Institute for Clinical Excellence (NICE), the third concerns what has traditionally been a rather opaque (and variable) aspect of clinical judgement (the decision to refer/operate) and the fourth addresses the potential (through linkages with other routine NHS data sets) for patient-assessed HRQoL to measure and monitor the equity consequences of policy decisions in the NHS. Finally, patient-assessed HRQoL could provide the data needed to tackle in part what has now been recognised as a significant problem in the NHS – the measurement of productivity.

Health technology (re)evaluation in the real world

The generally recognised “gold standard” method for evaluating the clinical effectiveness of healthcare technologies is the randomised control trial (RCT). However, while the RCT is the best method for isolating the impact of healthcare technology from other factors that may affect clinical outcomes, due to prohibitive limits on sample size and ethical considerations concerning patient inclusion in – or exclusion from – trials, RCTs do not reduce the uncertainty surrounding their results to zero.

Nevertheless, once a technology receives approval from NICE (largely, but not exclusively) on the basis of trial results, its actual use in the real world almost always differs from the controlled world of the RCT. For example, certain population groups such as women, children and the elderly are under-represented in clinical trials and yet are important users of healthcare (Harrison 2003; House of Commons Health Committee 1997). Harrison notes a study carried out in the Netherlands which found that more than a fifth of prescriptions for children were used “off-label” (that is, the use of a drug for a purpose other than that for which a licence had been granted). Grundy (2002) and Harrison and New (2002) have noted other similar examples concerning population groups, such as the elderly.

Clearly, the fact that significant numbers of people receive treatments that have not been tested on people like them is, as the Health Select Committee has noted, a surprising – if not astonishing – situation (House of Commons Health Committee 1997). However, in its role of assessing the clinical effectiveness of healthcare technologies, NICE recognises that, following its Final Assessment Determination (FAD), use of a technology in the NHS (or new trials) are likely to generate new data on effectiveness – and, indeed, safety – that will necessitate a review of the original FAD. When issuing original guidance, NICE also indicates a review date of anything from one to five years, though this is not fixed and is dependent on the emergence of new information (NICE 2004).

Apart from ad hoc opportunistic evaluations of the impact of technologies in their everyday settings and the standard mechanisms for the reporting of adverse drug reactions (ADRs – reported via the “yellow card scheme”²), there is no systematic gathering of patient-assessed HRQoL except in isolated settings (for example, see the BUPA programme, pp 26–7). Linked to specific treatments and patient and provider characteristics, such information would add an important dimension to the data NICE could consider in reviews of its guidance.

Patients’ values

That patients’ views about their healthcare are not just important but, in many cases, provide a vital input to a range of decisions is, we would suggest, incontestable. At present, the use of patients’ views in performance management is largely restricted to surveys of their satisfaction with the delivery of care – not its outcomes.

In carrying out its clinical and cost effectiveness appraisals, NICE states that evidence of effectiveness should include not only quantification of the effect of technologies on the course of a disease, but also the effect of technologies on patients’ HRQoL and the valuation of those effects in a manner that reflects the preferences of the general population. NICE also emphasises in its latest (revised) appraisal methodologies the importance of the patients’ (and carers’) perspectives on treatments. For example, NICE states that:

Patients and carers are a unique source of expert information about the personal impact of a disease and its treatment, which can help set the correct scope for the assessment of the evidence and enable the realistic interpretation of the clinical and economic evidence as the appraisal progresses. (NICE 2004)

Further, NICE notes that it is:

...interested in capturing a range of patient and carer views on, and experiences of, living with the condition, and the impact of technology on a patient’s symptoms and physical, social, psychological and emotional state, NICE (2004)

2

This scheme has been in existence for 40 years and, following a recent review, is, among other changes, likely to be extended to reporting of ADRs from patients directly.

And that, among other things, it is keen for appraisals to identify and measure:

...the impact of health technology on factors that matter most to patients, including physical or psychological symptoms, disability, function, long-term outlook, quality of life and lifestyle, (NICE 2004)

NICE has struggled to find a way of incorporating patients' views into the technology appraisal process (Devlin *et al* 2003). HRQoL data on health gain *as viewed by patients* creates the possibility that patients' views and values could be directly incorporated into health technology appraisals. Patients' views and valuations of their pre- and post-operative health states could have a role in informing assessments of the effectiveness and cost effectiveness of services (*for an extensive discussion, see Brazier et al 2004*).

Clinical thresholds

Medicine may be a science, but it is also an art, and there is ample evidence that clinical decisions as to when to start or to stop treatment and what form intervention should take varies, a variation only partly explained by clinical factors. Clinical opinion about, for example, what constitutes a reasonable waiting time, appears, in part, to be influenced by current waiting times (as well, we might presume, by the clinical condition of patients, their tolerance of waiting and other factors) (Appleby *et al* 2004). Understanding how clinical thresholds (the propensity to, for example, refer on to waiting lists or add to operating lists) vary and change in relation to the health status of patients would start to illuminate what has hitherto proved to be a rather opaque area of clinical decision-making, but with potentially significant policy implications.

Baseline measures of the HRQoL of patients being referred into the NHS, for example, would provide a ready means of gauging whether referrals are *systematically* changing in type as well as in number. For example, it is possible that patients with less severe conditions are being accepted for treatment – possibly as a result of changing referral behaviour as waiting times decrease. This would, in turn, provide a means of managing the flow of patients from primary to secondary care, and a means of checking that referrals are clinically appropriate. Baseline HRQoL measures could be used to develop, by clinical consensus, a “clinical threshold” on each instrument used – that is, a score on that measure below which there is general agreement that surgery will be of little or no benefit.

Pre-operative HRQoL assessments would also facilitate comparisons of the priority accorded to the treatment of patients with like characteristics between surgeons, trusts or geographical areas; that is, horizontal equity – the equal treatment (access to health services) of equals (in terms of health status). At present this remains obscured. HRQoL measures, taken at the point of referral, might also be used to generate a summary “score” that can be used explicitly to prioritise patients, in terms of their need and ability to

benefit from treatment, in a consistent manner across the NHS. HRQoL measures have been used for this purpose in New Zealand and elsewhere (Hadorn and Holmes 1997; Derrett *et al* 2002) to provide an explicit means of managing waiting lists.

The Department of Health has recently commissioned research to investigate the use of condition-specific and generic HRQoL measures to manage referral from primary to secondary care³. This illustrates the opportunities that exist for co-ordinating the selection of HRQoL instruments and the collection of HRQoL data.

Fairness and healthcare needs at the population level

Linked to other routine NHS data sets (for example, the patient record/hospital episode statistics – HES), HRQoL information could readily be used to monitor and measure differences in health between population sub-groups (for example, by age, sex, socio-economic group) that are the focus of health policy. The analysis of geographical differences in HRQoL, for example, could provide a more sophisticated means of proxying need in resource allocation formulae than is currently possible.

Measuring productivity

While extra spending is now addressing the historic under-investment in the NHS, it raises new questions: Where is the new money going? Is it being spent on the right things? Is our *health* improving as a result?

Knowing how much health the NHS produces may be the most crucial productivity question of all. However, with no routine monitoring of patients' health, we do not know how much extra health an extra pound spent on the NHS actually buys.

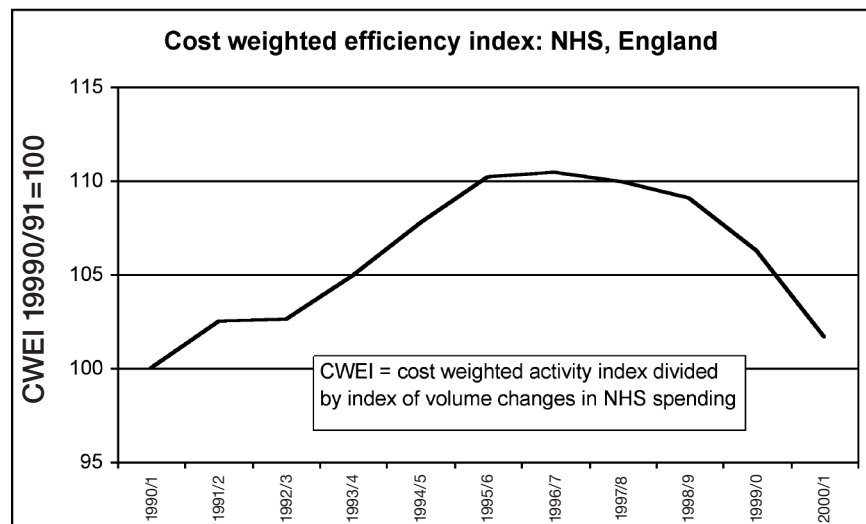
Traditionally, the NHS has tended to count the number of patients it treats, prescriptions written, operations performed, attendances at outpatient departments and other activities as measures of its output. Combined with the cost of producing these activities and divided by the extra money spent on the NHS each year, this gives a measure of NHS productivity or efficiency. If the NHS produces more activity each year than the increase in its financial inputs then productivity increases. But, as Fig 5 opposite shows, since 1996 productivity has been declining.

The traditional productivity measure is a ratio of outputs (activity) to inputs (money); the relatively large increases in NHS spending since 1997/8 have not been accompanied by similar increases in outputs. With spending rising even faster since 2000/01, this downward trend is likely to have continued in subsequent years. Extra spending has, in part, been absorbed by higher costs (rather than higher outputs); invested in services and activities, which may take some years to be reflected in increased outputs; increasingly channelled into activities not captured by the productivity measure; and used to increase the (unmeasured) *quality* rather than the (measured) *volume* of outputs.

3
The REFER Project (Realistic Effective Facilitation of Elective Referral), an assessment of the current use of referral tools for referral from primary to secondary care for elective surgical assessment and design of new tools – is being undertaken by researchers at LSHTM and Imperial College. Funded by the NHS Service Delivery and organisation research programme, it is an empirical study of priority-scoring tools and other systematic methods used to prioritise referrals from primary to secondary care for elective surgery.

Apart from these problems, there is another difficulty with activity-based measures of this kind – no matter how accurate and carefully constructed. It may seem somewhat paradoxical but it is not necessarily in the patient’s interests for the NHS always to do more. It is not, for example, inevitably desirable for the NHS to continually increase the number of admissions to its casualty departments; prevention is better. And as some drugs (and some operations and other interventions) are only of very limited benefit to patients it makes little sense for the NHS to strive to provide more.

Figure 5. NHS productivity: The cost-weighted efficiency index



Source: Adapted from: Public Expenditure on Health and Personal Social Services 2002 Memorandum received from the Department of Health containing Replies to a Written Questionnaire from the Committee HC 1210

Improving productivity is not just about producing more of everything for each extra pound, it is about doing the right things in the right way as efficiently as possible.

Addressing all these problems is not easy: while the Department of Health collects and collates information on changes in NHS costs, it has little data on the relationship between investment in prevention and eventual *outcomes*, or changes in the pattern of spending on activities not captured by its current productivity measure, nor changes in the quality of the services it provides. Research has recently been commissioned by the Department to investigate these issues and to address long-standing criticisms of the NHS productivity measure⁴. It will take some time before recommendations from this work will be ready to put into practice – as will other research into the general matter of public-service productivity (see Box 1, p9).

4 Patient Reported Health Outcomes Measures (PROM) in Treatment Centres

The measures that have been used to describe the productivity of the NHS as a whole are recognised as being seriously misleading. The Treasury and the Department of Health are now making a sustained effort to improve on them, fully taking into account the very difficult issues raised by any attempt to measure the performance of the NHS in a single figure. It will be some time, however, before this work produces useful results (see Box 2).

Box 2. Problems with measuring public-sector productivity

Recent rapid rises in public spending have increased interest in the issue of productivity and the more fundamental issue of measuring the outputs and outcomes of public services.

Indeed, although the UK is one of the few countries to follow internationally agreed best practice, the National Statistician recently announced a review of the future of government output measurement under the auspices of Sir Tony Atkinson, an expert in public economics.

Concerns about deficiencies in current methods of measuring output and productivity have also been raised by others, for example, the Treasury Select Committee (2003). In its latest report on the 2003 Pre-Budget Report, the Treasury Committee noted that: “A principal concern for the Government is the extent to which increases in public spending are feeding through into increased provision and quality of services combined with value for money, rather than higher costs.” (p 27, para 52).

The Committee heard from witnesses that there was a real issue about “measurement of outputs in the public sector; about capturing quality increases in outputs, about capturing the full range of outputs and about capturing investment in future capability”.

These criticisms are not new; measuring productivity – not just in the public sector but also in many (private) sectors of the economy – has always been problematic. In relation to the NHS, for example, attempts to quantify productivity and efficiency have been subject to criticism for many years (see, for example, Appleby and Little 1993; Appleby et al 1993; Clarke et al 1993; Appleby 1996).

There are clearly significant and difficult problems involved in measuring productivity in the NHS. However, on the issue of taking account of changes in the quality of services provided, patient-assessed HRQoL could have a role to play.

Conclusions

Shifting the NHS focus from the production of healthcare to the improvement of health is long overdue. HRQoL data has the potential to strengthen the management of performance of clinicians, surgical teams and hospitals. The data generated also has a wide range of applications throughout the NHS, by shifting the emphasis from process, activity and intermediate outputs to the outcome – patients' health. But is it feasible to use HRQoL data in this way? What are the challenges in measurement, and what are the costs and risks? We turn our attention to these issues in the following sections.

3

How should we measure changes in health?

There is no gold standard method for measuring health outcomes or health improvement. The “validity” of instruments used to measure HRQoL is often determined by comparison of the relative performance of one measurement approach vis-à-vis another. This complicates both the comparison of instruments (it is unlikely that one will be unequivocally “the best”) and the selection of appropriate measures that are fit for the purpose. Garratt *et al* (2002) noted the proliferation of measures of quality of life: in 2000, 1,275 separate measures were in existence, and considerable growth was evident in the production of new measures. In 1999 alone, 650 papers were published reporting the development or evaluation of such instruments.

Two main types of measures exist: those that are specific to a particular condition or disease; and those that are designed to capture patients’ experience of their health in a generic way⁵. Our discussion below is restricted to *patient-assessed* measures of HRQoL – that is, instruments designed for self-completion by patients – so patients’ views are put at the heart of the measurement approach.

Condition-specific measures

Condition-specific measures aim to capture aspects of health outcomes pertinent to specific health problems and improvements in those outcomes due to their treatment. A wide and ever expanding range of such instruments exists⁶, although some clinical areas (for example, rheumatology, musculoskeletal medicine, cancer, and older people) account for most of this research effort. Examples of condition-specific measures for four surgical areas are described in Table 2 (see p20).

The principal advantage of condition-specific measures is that they are generally sensitive to change in the conditions being described. The disadvantages of such measures in this context are that they restrict comparisons of health improvement and provider performance to providers of similar services or similar conditions. The improvement in patients’ health achieved by an ophthalmologist performing cataracts, for example, could not be compared to the improvement in health achieved by an orthopaedic surgeon. Even within a given clinical area – for example, trauma and orthopaedics, where there are a large number of surgical procedures – measures tend to be specific: for example, the Oxford Knee score is designed to capture health gain relating to knee surgery; performance on this measure could not be compared with the same clinician’s performance in undertaking hip replacements. Comparisons between providers’ performance are therefore more restricted. For some surgical interventions, such as hysterectomy, few condition-specific HRQoL measures are available, and those that are tend to focus on specific aspects of HRQoL, such as

⁵ There are also *dimension*-specific measures such as the Activities of Daily Living (ADL) that can be used on their own or alongside either generic or specific HRQoL measures

⁶ A searchable database of patient-assessed health measurement instruments is maintained by Oxford University: <http://phi.uhce.ox.ac.uk/phiidb.html>

sexual activity (Stead *et al* 1999); or have been developed and applied to specific indications for hysterectomy (for example, the Menorrhagia Outcomes Questionnaire captures outcomes following surgical treatment for menorrhagia due to benign disease – Lamping *et al* 1998).

Further, if the interest is in clinical performance *across* the NHS, there is no condition-specific measure that will be suitable for the diverse array of services offered by, for example, a general practitioner. Even in secondary care settings, many of the most common procedures undertaken in NHS trusts are of a somewhat non-specific nature. For example, among the top 20 procedures in 2003 are “invalid primary diagnosis”; “other admissions” and “planned procedure not carried out” (hospital episode statistics for 2002/3, reported by healthcare resource group). It is not obvious what condition-specific measure would be appropriate for such activities.

The large and increasing number of instruments makes the selection of a particular instrument difficult – those described in Table 2 overleaf were selected on the basis of their being widely used and accepted, following consultation with clinicians, or their merits as reported in the HRQoL literature.

In addition to measures associated with particular surgical interventions, many measures have also been developed to cover chronic diseases. In most societies chronic conditions are among the leading causes of death⁷, are the cause of the majority of the burden of poor health, disability and loss of quality of life, absorb a significant amount of healthcare resources, and tend (due to their chronic nature) to be well understood by patients, often better than by healthcare professionals (see Department of Health 2001). Measurement of the quality of life of people with chronic diseases is therefore particularly pertinent. Given this, it is perhaps surprising that some chronic disease areas, such as diabetes, still have no generally used, recommended or well tested measures (Bowling 2001). Table 3 (see p21) provides some examples of major chronic diseases and selected HRQoL measures.

7

The five leading chronic disease killers (chronic heart disease, all cancers, stroke, chronic obstructive pulmonary disease and diabetes) accounted for 64 per cent of all deaths in England and Wales in 2000.

Table 2. Condition-specific measures for health outcomes in four surgical areas

Surgical area:	Cataract	CABG	Hysterectomy	Trauma & Orthopaedics
<i>Selected instrument:</i>	Visual functioning (VF-14) (Steinberg et al 1994)	Integrated Therapeutics Group (ITG) CAD-specific short-form QoL. (Buchner et al 2001)	Menorrhagia Outcomes Questionnaire (Lamping et al 1998)	Western Ontario and McMaster University Osteoarthritis Index (WOMAC) (Bellamy et al 1988)
What is measured?	18 items that question a patient's ability to perform a variety of everyday visual tasks (for example, reading small print; recognising people); 14 contribute to the Visual Function Index	24-item CAD-specific questionnaire. Four scales: extent of chest pain; functioning and well-being; activities: physical; and activities: social	26-item questionnaire covering symptoms, post-operative complications; quality of life; and women's satisfaction with outcomes	Instrument designed specifically for the assessment of lower extremity pain and function in osteoarthritis of the knee or hip. Generates three scores: pain; stiffness; and physical function; and an overall score
Who completes the instrument?	Patients	Patients	Patients	Patients
Is there any indication of when and how measures should be taken pre- and post-surgery?	National Centre for Health Outcomes Development (2000): one week prior to surgery; 4 months post-surgery "to generally indicate the maximum improvement obtained" (p 38); 5 years later (to determine longer term benefit). There is an issue of when to take measurements of health outcome when one eye is operated on and then the other.	No	No	No
Is there evidence to support validity and reliability?	Yes (considerable literature exists)	Yes (Buchner et al 2001)	Yes – but only validated for hysterectomy patients suffering menorrhagia due to benign disease (Lamping et al 1998). Not validated for other indications for hysterectomy	Yes (considerable literature exists)
Where and how has the instrument been used in the UK?	Developed and tested in the US, it has now been validated and employed in the UK within the Cataract Outcome Study (Desai et al 1996). Used by BUPA to assess the performance of its ophthalmologists	Developed and tested in the US. Extent of use in the UK unknown	Developed and validated in England. Extent of use unknown	Developed and initially validated in Canada; validated and translated in a variety of countries and languages. Extent of use in the UK unknown

Table 3. A selection of condition-specific measures available for the measurement of health outcomes for selected chronic diseases

Back pain	Clinical Back Pain Questionnaire (Ruta et al 1994; Garratt and Ruta 1996). A 20-question, multi-item questionnaire measuring type, frequency, treatments and handicap stemming from back pain.
Arthritis	Arthritis Impact Measurement Scales (AIMS) (Meenan et al 1980; Meenan and Mason 1990). AIMS-2, a revised version of the original instrument, is a 78-item questionnaire. A short-form version (AIMS-2-SF) is available.
Epilepsy	Washington Psycho-social Seizure Inventory Scale (WPSI) (Dodrill 1978; Dodrill et al 1980). A 132-item scale (yes/no questions) covering 8 areas: family background, emotional adjustment, interpersonal adjustment, vocational adjustment, financial status, adjustment to seizures, management of medication, and patient satisfaction.
Asthma	Asthma Quality of Life Questionnaire (Juniper et al 1999). 152-items; responses recorded on a 7-point scale.
Schizophrenia	Schedule for Affective Disorders and Schizophrenia (SADS) (Endicott and Spitzer 1978). Semi-structured interview for clinical interviewer use.
Chronic heart failure	Chronic Heart Failure Questionnaire (CHQ) (Guyatt et al 1989). 16 items relating to physical and emotional functioning; items scored on a 5-point Likert scale.
Chronic depression	Centre for Epidemiological Studies Depression Scale (CES-D). (Radloff 1977). A 20-item, self-report scale. Covers symptoms of: depressed mood, feelings of guilt/worthlessness, sense of helplessness/hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance.
Cancer	Cancer Inventory of Problem Situations (CIPS) (Schag et al 1983). 131 problem statements grouped into more than 20 categories under 4 headings: personal care, medical situations, interpersonal interactions, miscellaneous.

Source: Bowling (2001)

Generic measures

Generic instruments, by contrast, seek to describe health and improvements in health in terms of the impact on health-related quality of life, broadly construed. The advantage of generic measures in this context is that they enable comparisons of health gain in a commensurate manner across all conditions, treatments, providers, specialist teams and hospitals. This generality can come at the expense of sensitivity: generic instruments sometimes fail to detect changes in conditions that might be regarded as “clinically important”. This is, for example, the basis for the common practice of incorporating both condition-specific and generic instruments in RCTs. While generic instruments may be unable sufficiently to detect HRQoL gains for some surgical interventions (for example, cataract removal – discussed further in Existing routine use of HRQoL in the UK, *see below*), for others (for example, CABG) there exists ample evidence that generic instruments are capable of capturing health improvements.

Examples of generic instruments include: SF-36® (and concatenated versions such as the SF-12®); the Nottingham Health Profile; Sickness Impact Profile; the QWB; AQoL; WHOQoL; and EQ-5D (Bowling 1997; Drummond *et al* 1997; Essink-Bot *et al* 1997). All these instruments provide a means of *describing* (generating a “profile” of) health. Instruments that restrict their description of health outcomes to the reporting of multiple sub-scale scores are of limited use, since they fail to indicate the overall direction and magnitude of change as assessed by the patient. Thus an important consideration in the selection of an instrument for the purpose of performance management is that it is also capable of generating an aggregate measure or summary score, representing the patients’ *overall* assessment of their health. The SF-36® applies an algorithm to patients’ responses to individual questionnaire items and scales to produce two summary scores: one for physical health and one for mental health. The EQ-5D uses a visual analogue scale to elicit from patients a single score for their overall health. Of course, the information from both these instruments are capable of being presented in a disaggregated form. The characteristics of these two instruments are summarised in Table 4 opposite; more detail on the means by which each describes health are presented in Appendices A1 and A2.

Existing routine use of HRQoL in the UK

While, as we noted in section 1, the use of HRQoL measures has not been prominent in the NHS, there are nevertheless a number of researcher-led routine (that is, excluding time-limited trials and RCT-type studies) uses of HRQoL in the NHS and the UK private sector that may provide valuable data and experience.

The Health Outcomes Data Repository⁸ (HODaR) contains EQ-5D and SF-36® data that have been routinely collected from patients at a large UK NHS trust, and which are linked to clinical data. The database is available commercially and currently contains

8

www.hodar.co.uk

Table 4. Characteristics of two generic health state measures: SF-36[®] and EQ-5D

Characteristics:	SF-36 [®]	EQ-5D
How is health described?	A 36-item questionnaire. (Sample page from the SF-36 [®] is available at: <a href="http://www.SF-36<sup>®</sup>.org/tools/pdf/SF-36<sup>®</sup>v2_Standard_Sample.pdf">http://www.SF-36[®].org/tools/pdf/SF-36[®]v2_Standard_Sample.pdf)	5 dimensions (mobility; self-care; usual activities; pain/discomfort; anxiety/depression); 3 levels within each (no, some or extreme problems). (2-page questionnaire included in Appendix 1)
Does it generate a summary score?	Scores can be presented as Physical Component Summary Scores and Mental Component Summary Scores, each ranging from 0 to 100. SF-36 [®] does not generate a single “summary” index	The descriptive system is accompanied by a self-rated 0–100 score, summarising the patient’s assessment of their own health.
What is the evidence base?	Extensively researched and widely used. Subject to ongoing research and development	Developed and promulgated by the EuroQol Group. Subject to ongoing research and development.
Parsimony/responder burden?	The SF-36 [®] is suggested to take 5–10 minutes to complete. Shorter versions, the SF-12 [®] and SF-8 [®] , are available	Takes 1–2 minutes to complete; comprises 2 pages.
Access/public domain?	<a href="http://www.SF-36<sup>®</sup>.org/">http://www.SF-36[®].org/ Users of the SF-36 [®] must register their use and obtain a user licence from the Medical Outcomes Trust (MOT), the Health Assessment Lab and QualityMetric Incorporated	http://www.euroqol.org/ EQ-5D instruments are available in the public domain and are freely available. Users are encouraged to register their use with the EuroQol Group; manuals may be purchased
Use in the UK?	Widely used in clinical and population health research in the UK and internationally. Adopted by BUPA for measuring the performance of healthcare providers	Wide use in clinical studies, patient groups and general population studies. Population norms are available. No previous use in measuring the performance of clinicians that we are aware of.

data from more than 30,000 patients. The data is principally used by pharmaceutical companies for various purposes such as epidemiology, burden of disease and modelling studies. Pre-admission data are not collected – questionnaires are sent to patients six weeks following discharge. The consultant treating the patient is not currently recorded, though this could be linked to the database by the hospital (private correspondence, Simon Dixon, 12 May 2004). The EQ-5D and SF-36® have both been used in large surveys of the general public – population norms are available for both that may have relevance in “benchmarking” performance.

The Picker Institute (Europe) has, since 2001/2, been developing and conducting national surveys of NHS patients to assess the quality of patients’ experience of care in acute trusts. The 2001/2 inpatient survey incorporated the EQ-5D profile (but not the visual analogue scale overall assessment of health), although this has subsequently been dropped⁹, and the most recent survey includes only a single health status question¹⁰.

The most valuable evidence on the potential use of HRQoL in performance management comes from BUPA’s pioneering use of the SF-36® (BUPA 2003). Since 2002, questionnaires have been routinely sent to all patients prior to admission and follow-up questionnaires are sent 12 weeks following surgery. The response rate achieved so far for the baseline assessment across all BUPA hospitals is, on average, 10–15 per cent, but much higher in some individual hospitals. Incentives to boost completion rates are currently increasing responses. Follow-up response rates are currently around 75 per cent. Following concern over the performance of the SF-36® with respect to cataracts, this was dropped in favour of a visual function-specific measure, the VF-14 originally designed for cataract but now validated for other conditions. Data from these patient surveys have enabled BUPA effectively to identify “outliers” in terms of both good and poor performance – and to feed that data back to managers in a format that has enabled quality issues and concerns regarding individual consultants to be addressed. We will refer to specific aspects of BUPA’s experience throughout the remainder of this report.

When should HRQoL be measured?

The routine use of HRQoL requires a decision about the appropriate interval between baseline and post-treatment measurement. This interval would need to be consistent between the patients treated by each clinician, clinical team and hospital if comparisons of changes in HRQoL are to be a valid basis for inferences about their relative performance.

As noted above, HODaR has opted for a six-week interval; and BUPA for 12 and 16 weeks for the SF-36® and VF-14 respectively. Before adopting its current protocol,

9
Due, in part, to a view that overall health could be captured by the single health question that was retained

10
www.nhssurveys.org/doc/Amendments_inpatient_survey.pdf

BUPA piloted SF-36® measurement at six weeks, 12 weeks and one year. Measurement at 12 weeks was considered better to capture health-related changes in quality of life following recuperation from surgery than six weeks; longer-term HRQoL at one year was considered relevant, but response rates were lower (personal correspondence, Andrew Vallance-Owen, 18 May 2004).

Given the diverse procedures performed in the NHS, imposing a standard measurement interval across all procedures may not be valid. BUPA, for example, has found that it is not so useful for diagnostic procedures such as scopes. Differing conditions and procedures will generate different time profiles of benefit or disbenefit from treatment. A given interval – say, three months – may capture 90 per cent of the eventual change in HRQoL for some patients, but 50 per cent of the change in others. However, where the objective of measurement is the detection of *variation* in the production of HRQoL gain, and “outliers” in performance, this issue may be of less importance than if the (or an) objective is to measure HRQoL gain *per se*.

The choice of interval is more complicated for procedures such as cataract removal and bilateral hip replacements, where the post-surgical HRQoL measurement might be taken either as the standard interval following the first surgery, or as the standard interval once *both* eyes/hips have been operated on.

Finally, the foregoing presumes a clear-cut difference between pre- and post-treatment periods and assumes a bias towards acute (surgical) hospital care. The ongoing care and treatment/symptom relief associated with many interventions for chronic diseases, for example, will require a different approach to the monitoring of patient-assessed HRQoL, although there are tools (such as the MDS/RAI¹¹ being used in BUPA care homes) that may be useful when administered sequentially over time.

How should the data be reported?

We noted in Generic measures (see p22) that HRQoL measures that produce a single score summarising the patients’ overall health have an important advantage over measures which produce only profiles or multiple sub-scores that cannot be aggregated. This advantage lies both in the ability to compare overall performance (where sub-scores may move in conflicting directions) and in the parsimony of reporting key information for management purposes. Routine use of the instruments described in Condition-specific measures (see p18) and Generic measures (see p22) would provide a rich and complex source of data on patients’ experience of their health before and following treatment. These data can be used to explore a wide range of questions and issues. However, their effective use in performance management requires not complexity but the clear communication of relevant variations, both to clinicians and to managers, in a manner that can be acted upon.

¹¹ Minimum Data Set (MDS)/ Residential Assessment Instrument (RAI). See, for example, <http://www.cms.hhs.gov/medicaid/mds20/mdsdates.asp>

There is clearly a trade-off here, with potentially useful information being lost in the summary scores – for example, either on the EQ-5D's visual analogue score, or the SF-36®'s algorithm-deduced scores for physical functioning and mental health.

However, BUPA's experience in reporting the SF-36® data was that presenting results in terms of performance on sub-scale components was often misinterpreted; and that much of the variation was, in any case, not significant. BUPA opted to present its results in terms of "Shewart" charts (see Fig 6 opposite), which focus on the identification of outliers – those generating patient outcomes more than two standard deviations above or below the average change in patients' HRQoL¹². Reports to hospital units and consultants focus on placing their performance in the context of the mean, range and distribution of performance.

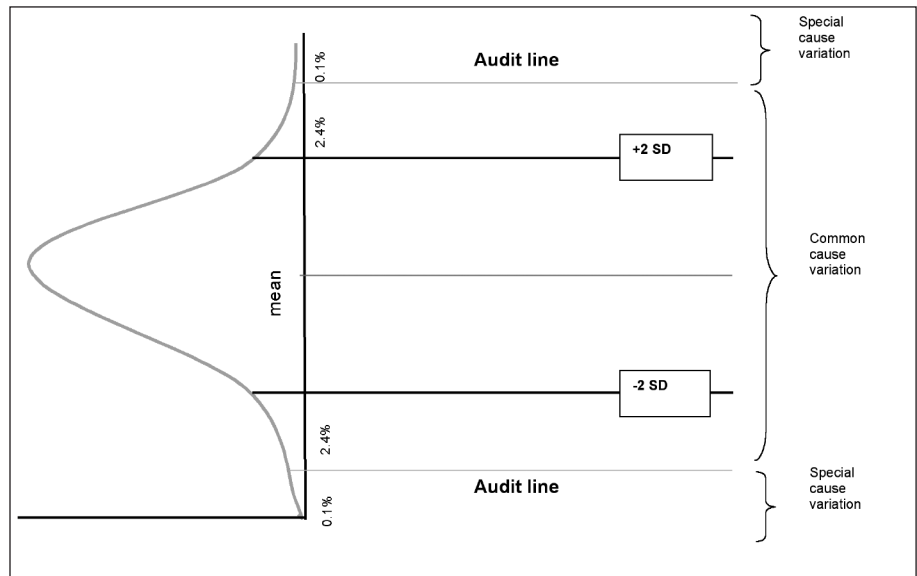
Conclusions

Generic measures offer considerable advantages in assessing and comparing HRQoL change across the diverse range of activities and procedures undertaken in the NHS – but the principal concern is their sensitivity in detecting changes in health resulting from specific conditions and interventions. Choice of instrument should be informed both by existing use and evidence (informed by data and experience that are already available) and by relevant developments in the NHS. For example, there are a number of initiatives under way or planned in the NHS that involve the use of HRQoL measures. These include investigations of their use in managing referrals (the Department of Health-funded REFER project now under way); and their use in managing the performance of treatment centres (the PROM research programme currently being commissioned by the Department). There are important opportunities for the selection and use of HRQoL to proceed in a coherent and co-ordinated manner across these various activities.

¹²

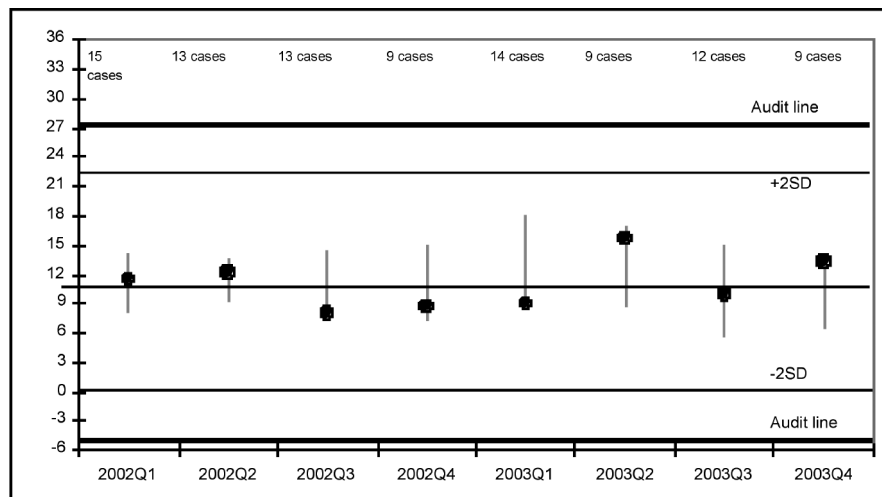
In large populations, of course, this means identifying around one in 20 as outliers, and so more stringent cut-off points (for example, three standard deviations from the mean) may sometimes be desirable.

Figure 6. Identification of outliers: BUPA's presentation of HRQoL data



Source: BUPA (2003)

Figure 7. Identifying the performance of individual providers:
Primary hip replacement: Changes in SF-36 physical summary score



Source: BUPA (2003)

4

What are the risks and costs?

The collection, collation, analysis and activities associated with the generation of patient assessed HRQoL and its consequent benefits outlined in section three will not, of course, be costless. Apart from the direct financial costs of administration, the routine collection of HRQoL data will necessarily carry risks and imply challenges – not least in relation to the way such data is used to actually improve healthcare performance. In this section we identify cost issues as well as potential difficulties and concerns such as gaming, the need for clinical buy-in and possible problems with the use of HRQoL data.

How much will it cost?

The instruments considered earlier in Tables 2 (see p20), 3 (see p21) and 4 (see p23) are *patient assessed* – they involve the patient completing a hard copy questionnaire. Most of these instruments are very brief indeed – two pages – take just a few minutes to complete, and are designed to be easy to use and suitable for (unaided) self-completion.

As such, they could be administered without the requirement of nursing, consultant or clerical assistance. Indeed, to help avoid gaming and to improve accuracy it is preferable if the instruments are completed outside the surgical context; for example, in the GP surgery or the patient's own home, rather than at the outpatient appointment; and follow-up measures posted rather than completed in a post-surgery checkup. In the most simple scenario, at the point of referral, the patient could be asked to complete such an instrument before they leave the GP surgery (or when back in their own home); post-operation, patients' records could be used to flag the mail-out of a questionnaire at some chosen interval following surgery (for example, at six months).

While the actual direct unit cost of administering HRQoL instruments will be relatively small costs of implementing a collection and reporting system are likely to be relatively large and contingent on decisions about the following issues.

Language

Most of the condition-specific instruments referred to are validated in English language formats. Both generic measures – EQ-5D and SF-36® – have validated translations of their instruments available in a range of languages. Some of the condition-specific instruments are also available in multiple languages – for example, WOMAC is available in Swedish, Hebrew, Korean and German.

If an instrument is chosen for which appropriate alternative language versions are *not* available, this risks either (i) a biased response rate (non-English speakers will be excluded from the process) or (ii) more time-consuming administration, via an

interpreter or assistance from a health professional. This has not been an issue for BUPA: although it has yet to analyse the sociodemographic characteristics of non-responders to its surveys, the socioeconomic and ethnic composition of its clients is likely to be less diverse than that of the NHS – although this depends on the “source” of patients (that is, whether personally insured, covered by company group insurance or paid for by the NHS). Avoiding these biases will be both more important and more challenging in the NHS.

Method of administration

Hard copy questionnaires are the most common format for this sort of exercise. There are alternatives – such as the use of on-line software or patient in-home monitoring devices¹³ – but these require the equipment itself and training in its use, or access to computers. Alternatively, some of the instruments are feasible to administer via brief telephone interviews (for example, Derrett *et al* (2002) used both condition specific and EQ-5D instruments in this manner). This may confer some advantages in terms of completion rates but is more time-consuming to administer. If paper and pen is the preferred mode of administration, costs (but also the response rate) will be higher if a follow-up reminder letter and duplicate questionnaire are sent to non-responders.

Decisions concerning the electronic storage of HRQoL data, their link to patient records and how such data would fit with plans for the electronic patient record (EPR) and in particular the NHS National Programme for IT (NPfIT) will have a significant impact on costs. There is, as far as we understand, no provision in NPfIT for accommodating the routine collection, processing and other administrative requirements associated with HRQoL information.

Sample size and coverage requirements

The options in terms of the comprehensiveness of data collection are either to carry out a census, including all patients in such an exercise, or to take a random sample (for example, one in 20). The issue is: what is the *smallest* sample that would be required in order to detect important between-clinician variation? BUPA, for example, sought SF-36[®] measures for *all* patients – the more costly option. While it might not be necessary to survey all patients if the objective is to determine health gain from a given procedure, where the objective is to examine variation in the *delivery* of this procedure by clinicians, it may be more appropriate, initially at least, to include all patients. Note that, in practice, data on all patients will not be achieved because the response rate to the baseline and post-surgery survey will be less than 100 per cent.

In addition to the choice of census or sample, there is also a decision concerning coverage in terms of health care intervention/treatment. As noted earlier, there is an implicit bias towards the use of patient-assessed HRQoL for one-off, usually surgical, interventions with a fairly clear-cut difference between pre- and post-treatment

13

For example, the REALITY project is currently evaluating the use of hand-held monitoring devices, operated by patients, to collect *daily* information on clinical symptoms for patients with a range of chronic conditions, and is experimenting with collecting condition specific and generic HRQoL data using this same technology

periods (and health states). However, the majority of patient contacts with the NHS do not fit this description – for example, the management of many chronic diseases (which may, in fact, involve the controlled management of deterioration rather than improvement in health state), baby development check-ups by health visitors, the (often social) care provided by district nurses to elderly people, and visits to general practitioners involving reassurance or “watchful waiting” but no direct treatment. Clearly, collecting patient-assessed HRQoL for every GP attendance, for example, would add hugely to costs.

Will variation in performance be detected?

It is important that the selected HRQoL instruments are “fit for purpose” – that they can adequately (and unequivocally) discriminate between “good” and “poor” providers of healthcare. Derrett *et al* (2003) show that post-operative health gain, measured on the SF-12®, the EQ-5D and a range of condition-specific measures, was poorly correlated between measures. This suggests that how much health gain is captured may depend very much on which instrument is used. A number of health outcomes researchers we consulted expressed concerns about whether HRQoL measures would be sufficiently sensitive to capture differences between individual providers. Others considered the simple “before and after” comparison to be too simplistic, cautioning that individual differences in the patient experience, and “noise” in their self-reported health may not necessarily “wash out.” We note this concern – but also note that the primary intention is not to measure health gain from treatment, but the *variation* in health gain across individual providers. BUPA's experience suggests the SF-36® was capable of performing this function across a range of surgical procedures, with the exception of cataract removal.

Further, research is required to determine whether the effect of the clinician (or provider) is distinguishable from *other* determinants of health gain following treatment. If measures are taken at baseline and after intervention, can change in these measures confidently be attributed solely to the clinician, for example, rather than to the team involved in the patients' clinical pathway? Further, variations in aspects of follow-up care outside the clinicians'– or treatment centres' – control, such as home support, primary care, domiciliary services and prescribing, are additional potential factors that may confound interpretation of variations in HRQoL between providers.

However, it is worth noting here that there is probably much useful information to be gained from multiple administrations of a health questionnaire at various points in a patient's care pathway in order to capture contributions (or, at a minimum, changes) to a patients's health gain (or loss) from the different stages in treatment and care – from surgery through to rehabilitation and physiotherapy, for example.

Importantly, however, there is the question of confounding as a result of variations in the casemix of patients between clinicians/hospitals. The degree to which, say, change in HRQoL is dependent on patients' age (for example, older patients may have less capacity for gain than younger patients), pre-existing conditions and overall health status prior to intervention is an empirical issue and one that will vary in importance across interventions. However, while not underestimating the difficulties involved, the problem of comparability/interpretability of variations in HRQoL are not insurmountable; given the data on important casemix factors, statistical techniques can provide the tools to adjust HRQoL results to isolate "real" variations (cf Crosby *et al* 2003; Ireson *et al* 2002; Dennison 2002).

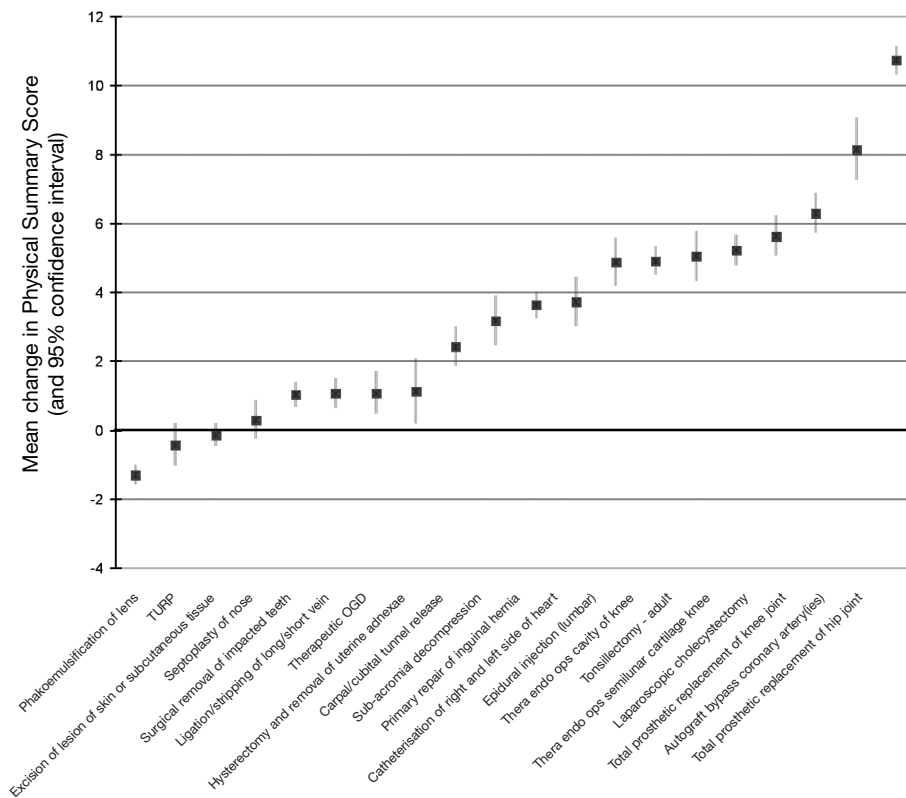
Clinician support

Clinician buy-in will be important in ensuring both compliance in the collection of data and its usefulness in measuring and managing performance. Some of those we consulted in preparing this report indicated concern about basing the management of clinicians' performance on measures that are "purely subjective", that is, patients' assessments of health gain, whether condition specific or generic, rather than clinical indicators that are seen to be more "objective" and reliable (and possibly more familiar). Opinion – and experience in the use of such instruments – varies considerably across different clinical areas.

However, opposition may not be as great as might be imagined. The Royal College of Psychiatrists, for example, has indicated that it is "broadly supportive of the introduction of routine outcome measurement" (personal correspondence, 27/5/04). Correspondence from the Royal College of Ophthalmologists included the comment that the routine collection of HRQoL in the NHS was "not only a good idea but inevitable" (18/5/04). In the process of talking to specialists on these issues, we discovered that in some specialties – for example, orthopaedics – there is already widespread, routine use of both generic (most commonly the SF-36®) and condition specific (most commonly WOMAC or Oxford Hip and Knee scores) (personal correspondence, Chris Moran, 27/5/04). These uses are clinician initiated and, although the data may be reported in the clinical literature, they are not used to identify the performance of individual clinicians, nor are they linked to managerial processes to improve quality and performance more generally. Nevertheless, this suggests a groundswell of support among some clinicians that HRQoL measures provide useful information that can assist their medical practice. Consultants working in both the NHS and the private sector will already be familiar with the use of routine HRQoL measurement by BUPA. Consultant level comparisons using HES data are already being piloted by the Department of Health (Royal College of Physicians 2004).

BUPA's experience indicates that health gain on the SF-36® varied considerably by type of procedure (see Fig 8 below), with some procedures appearing to generate substantial improvements in health, while others resulted either in no change or negative change within the measurement period. This might be expected to have implications for the acceptability of the process to surgeons within each specialty – and was, in BUPA's case, the principal reason for its shift to the use of the condition specific VF-14 for cataracts, which was better able to detect change in HRQoL.

Figure 8. Changes in the SF-36® physical summary score after 12 weeks, by surgical intervention



Source: BUPA (2004)

The manner in which information is reported back and used is likely to be of key importance to the acceptability to clinicians. BUPA, for example, provides quarterly reports to hospital managers on the performance of the surgical *team*; only the consultants themselves receive reports on their individual performance. Respecting the anonymity of individual clinicians carries a trade-off of “blunting” the use of the data to pinpoint and manage unacceptable performance. A solution is to confine reports to surgical teams, but for those reports to provide an indication to managers if there is an individual (without identifying them) whose performance is an outlier in terms of the team or in terms of the mean for that procedure more generally.

Patient-reported outcome information does not, of course, pinpoint the causes or underlying reasons for variation; HRQoL information provides providers with an important flag that there may well be something worth further investigation and audit. How this is pursued is usually the main concern for individual clinicians and teams.

Gaming

The potential for gaming by stakeholders – patients, doctors or hospitals – needs to be considered both in data collection and performance management. For example, if clinicians have a degree of control over the selection of patients from hospital waiting lists, they may attempt to select patients whose gain from treatment is greatest. Likewise, if patients consider that the baseline description of their health that they provide may have some bearing on the decision to offer them treatment – and how quickly that may take place – there may be an incentive for them to over-report the impact of their condition on their (baseline) HRQoL.

Will the information be used? Will it result in improvement?

In Clinician support (see p31), we outlined some options for the way in which data could be reported. To *whom* these measures of performance are reported, and how they are acted on, is crucial to whether HRQoL data makes a difference to the performance of health services. Marshall *et al* (2000), examining the use and impact of information on the performance of hospitals and health professionals in the US, note:

Consumers and purchasers rarely seek out the information [on physician/hospital performance] and do not understand or trust it; it has a small, although increasing, impact on decision-making. Physicians are sceptical about such data and only a small proportion makes use of it. Hospitals appear to be most responsive to the data. In a limited number of studies, the publication of performance data has been associated with an improvement in health outcomes.

Evidence to support the claim that patients and purchasers actively use performance evidence is weak (Hibbard 2003), as is evidence on the link between evidence on performance and change in performance. Leatherman *et al* (2003) note:

A curious aspect of the field of quality measurement and reporting is the relative dearth of objective measurement of its impact. In other words, objective evaluation with a feedback loop for learning, which is a fundamental tenet of quality improvement, is insufficient in its self-application. Much of the “evidence” of the effectiveness of quality measurement – its alleged power to lead to quality improvement – is anecdotal narrative. (p83)

Mannion and Goddard (2001), investigating the impact of publication of clinical outcomes data on NHS Trusts in Scotland, found that the indicators

...had a low profile in the trusts and were rarely cited as informing internal quality improvement or used externally to identify best practice. The indicators were mainly used to support applications for more funding and service development.

The poor effect of clinical outcomes data was attributed to: a lack of professional belief in the indicators, arising from perceived problems with data quality and time lag between collection and feedback; limited dissemination; weak incentives to take action; a “predilection for process rather than outcome indicators”; and a belief that informal information is of more use than quantitative data in assessing clinical performance. Identifying ways of overcoming these problems will be key to the success of using HRQoL data.

Much of the existing performance measurement in the UK is focused on adverse outcomes – death and readmission rates. Measuring performance in terms of gains in health facilitates a focus on the positive as well as the negative outcomes. Deviations from the average in terms of excellent performance offers an opportunity to identify, reward and disseminate especially good practice. Focusing managerial and clinical attention on “good outliers” is important, and offers scope for continually increasing overall quality and performance.

Nevertheless, the identification of “bad outliers” is of importance for other reasons, including the responsibility of the NHS to its patients. In most cases, it would be hoped that these problems would be detected and addressed through the hospital’s management. Evidence of persistent (that is, in more than one period) poor performance (for example, a clinician with patient outcomes more than two/three standard deviations below the mean – although “underperformance”¹⁴ may also be of concern) may be addressed in a number of ways; for example:

¹⁴ That is, consistently poor but not an outlier as defined earlier.

- This evidence could be considered during the annual appraisal of clinical staff, and would be relevant to the revalidation process commencing in April 2005
- The GMC may consider acting on such evidence in informal cautions or formal procedures
- The Healthcare Commission could utilise such data in its regular provider reports of broader assessments of standards of clinical care

Finally, performance in terms of HRQoL is likely to complement, rather than replace, existing ways of measuring performance. For example, managers are likely to be interested in improvements to both health *and* activity (measured in terms of Finished Consultant Episodes (FCEs) per consultant). The expected relationship between these two dimensions of performance is not obvious. Similarly, the relationship between health outcomes and *other* measures of performance for hospitals (including patient satisfaction with standards of care, waiting times and other targets or indicators of quality of care – see Mant and Hicks 1995) may be complex and remains largely unknown.¹⁵

Conclusions

There is some uncertainty over how best to detect and report unacceptable variation in performance. The risks and costs described here suggest that more information is required before large-scale implementation can confidently be recommended for the NHS. It is our view that questions about how and when to measure HRQoL are certainly not insurmountable, and a pilot study – in conjunction with analysis of existing HRQoL data collection efforts – can address these. However, the crucial question is whether the NHS is prepared to make a real commitment to improving performance, backed up by appropriate actions. Unless measurement is accompanied by rewards for good performers, and an action plan for poor performance, there is little to be gained from introducing yet more measurement.

15

Techniques such as Data Envelopment Analysis (DEA) would provide a means of analysing the interaction between multiple facets of performance, such as the quantity of outputs and quality of outcomes, and identifying and benchmarking best performance.

5

What next?

There is potentially much to be gained from routine before-and-after measurements of health, both as a means of monitoring and managing the performance of providers and as a means of facilitating a system-wide refocusing of the NHS on health. However, collecting data is a means to an end, not an end in itself, and the NHS needs to be very clear about how it would use such information in practice if these gains are to be realised.

Further, there remain some important practical questions:

- Which HRQoL measures best discriminate between good and poor providers?
- How often, when, and how much would it cost to collect and analyse these data?
- How can the information be presented so that it generates action – and what improvements will result?

More information is required before a routine use of HRQoL can be advocated. The bottom line in considering routine use of HRQoL to manage the performance of NHS providers is not only whether this will make a positive difference to performance and the well-being of patients, but also the size of this difference. In short, will the benefits outweigh the costs involved?

In this section, we provide our recommendations about how to proceed. In essence, these are: first, to make sure existing data are fully analysed; second, the need to establish a pilot study to answer remaining questions; and third, given that routinely collected HRQoL data could be used in many ways and that there are current examples of HRQoL collection, to ensure that opportunities to “connect up” these activities are not missed.

Analysing existing data

Given the cautions offered in Will the information be used? Will it result in improvement (see p33), evidence is required on how measuring HRQoL impacts on *changes* in quality. BUPA’s innovative use of the SF-36[®] is commendable, and has clearly demonstrated the feasibility of using such data in performance management. However, what at present remains unknown about BUPA’s experience is whether it can be successfully transposed from the private sector to the NHS and what demonstrable change it has generated. Evidence of success might be expected to be apparent in (at least) two ways:

- **The standard deviation should decrease:** that is, variations in provider performance should decrease, and the shape of the distribution – particularly in the “tail” of poor performers – should change.
- **The mean or average level of performance should increase,** as both poor and good quality performers respond to incentives.

Further analysis of BUPA’s data could offer important – and perhaps unique – insights into these questions. Also, given relatively low response rates to the baseline assessment reported by BUPA, there is an important question of how to ensure adequate response rates appropriate to subsequent analysis and action.

More generally, we have noted throughout our consultation with clinicians and others, the number of clinicians who are, on their own initiative, already using condition specific and/or generic instruments routinely in their practices. These are usually motivated by research objectives rather than quality assurance or performance measurement *per se*. Sometimes these data are coded to identify the doctor or surgical unit involved. Because the focus is on the effectiveness of *treatment* – not the effectiveness of the provider *delivering* the treatment – these data have generally not been analysed to identify between-clinician variations. There is some scope for analysing existing data sets where the provider has been coded (or encouraging research teams to incorporate provider codes) to explore the extent of variation in HRQoL change between providers on the instruments.

Recommendations for a pilot study

Given the uncertainties associated with the routine collection and use of patient-assessed HRQoL in the NHS, there is a clear need for a pilot study. This should be designed to investigate the feasibility, benefits and costs of using HRQoL to manage providers’ performance. Specifically, the pilot study should be designed to provide evidence on:

- The feasibility of data collection (for example, response rates) and acceptability to patients of data collection (for example, in terms of ease of completion; the availability of appropriate language versions; the availability of assistance where the patients are unsure how to complete the forms and so on).
- The way in which patients can, and do in practice, make use of the results to inform their choices between providers.
- Acceptability and usefulness to doctors.
- Precise cost projections for the collection and analysis of the data if rolled out across the NHS/expanded across services/interventions.
- The feasibility of taking a sampling approach (for example, one in 20) versus a census.

- The relative performance of condition-specific and generic measures in detecting variation in performance between doctors, surgical teams and hospitals.
- The best timing and frequency of seeking health outcomes measurements from patients. Options include: pre-operative measurement at the point of referral or outpatient consultation; measurement during treatment; post-operative measurement taken at intervals following treatment (and variations in the appropriate interval between therapeutic areas).
- How the information can be or is used to identify poor performers – and the most effective means of communicating this to doctors, hospitals and managers.
- How the information is used (for example, in annual appraisal processes) and what resultant actions are taken.
- The relationship between poor performance on health outcomes and other measures of provider performance.

Co-ordination of HRQoL initiative across the NHS

This report has indicated a number of ways in which routinely collected HRQoL data could improve the performance of the NHS and, most importantly, the health of patients. Indeed, a wide range of individual initiatives involving HRQoL are already under way in the NHS, from attempts to incorporate patients' views into NICE technical appraisals to management of referrals from primary to secondary care, and the use of HRQoL to assess the performance of treatment centres. There are important opportunities for the co-ordination of selection and use of HRQoL instruments – in a manner that will avoid duplication of effort and will maximise system wide benefits – that should not be missed.

6

Conclusions: Time for action

Shifting the NHS focus from the production of healthcare to the improvement of health is long overdue. HRQoL data has the potential to strengthen the management of performance of clinicians, surgical teams and hospitals. The data generated also has a wide range of applications throughout the NHS, by shifting the emphasis from process, activity and intermediate outputs to the outcome – patients' health. But there are choices to be made about how to measure outcomes and how to use the data generated.

There are many disease, treatment and patient-group specific HRQoL measures. However, generic measures such as the SF-36® and the EQ-5D offer considerable advantages in assessing and comparing changes in patient-assessed HRQoL across the diverse range of activities and procedures undertaken in the NHS – but the principal concern is their sensitivity in detecting changes in health resulting from specific conditions and interventions. Choice of measurement tool should be informed both by existing use and evidence and by relevant developments in the NHS such as the Department of Health-funded REFER project now under way and the PROM research programme currently being commissioned by the Department to investigate ways of routinely assessing the performance of treatment centres.

There is some uncertainty over how best to detect and report unacceptable variation in performance. The risks and costs described in section 4 suggest that more information is required before large-scale implementation can confidently be recommended for the NHS. It is our view that questions about how and when to measure HRQoL are certainly not insurmountable, and a pilot study – in conjunction with analysis of existing HRQoL data collection efforts – can address these. However, the crucial question is whether the NHS is prepared to make a real commitment to improving performance, backed up by appropriate actions. Unless measurement is accompanied by rewards for good performers, and an action plan for poor performance, there is little to be gained from introducing yet more measurement.

Overall, having reviewed evidence of the potential benefits and costs of routine measurement of outcomes, current knowledge and experience of how to produce this information, and current clinical and technical opinion on this matter, more information is required before the routine use of HRQoL can be advocated. The bottom line in considering routine use of HRQoL to manage the performance of NHS providers is not only whether this will make a positive difference to performance and the well-being of patients, but also the scale of this difference. In short, will the benefits outweigh the costs involved?

Three issues need further investigation:

- Which HRQoL measures best discriminate between good and poor providers?
- How often, when, and how much would it cost to collect and analyse these data?
- How can the information be presented so that it generates action – and what improvements will result?

Appendices

A1 The EQ-5D

A2 The SF-36[®] questionnaire

A1. The EQ-5D descriptive system

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today

Mobility

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

Self-Care

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

Usual Activities (e.g. work, study, housework, family or leisure activities)

- I have **no** problems with performing my usual activities
- I have **some** problems with performing my usual activities
- I am **unable to perform my usual activities**

Pain/Discomfort

- I have **no** pain or discomfort
- I have **moderate** pain or discomfort
- I have **extreme** pain or discomfort

Anxiety/Depression

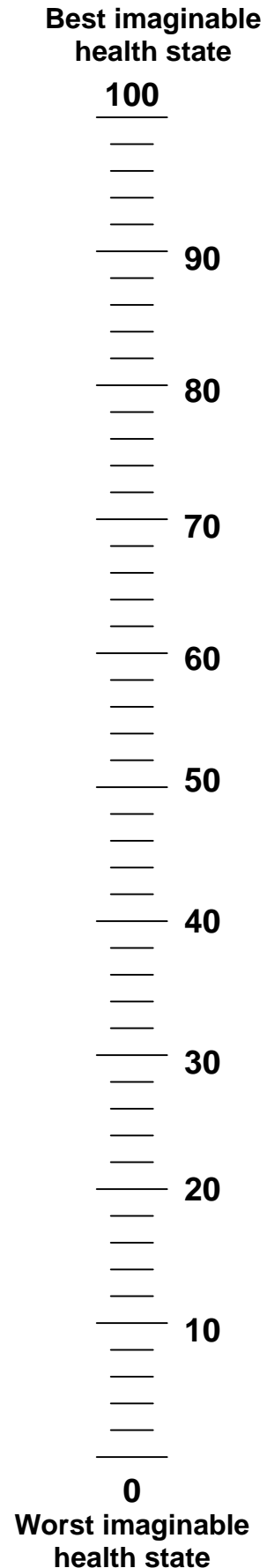
- I am **not** anxious or depressed
- I am **moderately** anxious or depressed
- I am **extremely** anxious or depressed

The EQ VAS

To help people say how good or bad their health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is on this day, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

**Your own
health state
today**



A2. The standard SF-36® questionnaire

Your Health and Well-Being

This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities. *Thank you for completing this survey!*

For each of the following questions, please mark an in the one box that best describes your answer.

1. In general, would you say your health is:

Excellent	Very Good	Good	Fair	Poor
▼	▼	▼	▼	▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Compared to one year ago, how would you rate your health in general now?

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
▼	▼	▼	▼	▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	No, not limited at all
	▼	▼	▼
a <u>Vigorous activities</u> , such as running, lifting heavy objects, participating in strenuous sports	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b <u>Moderate activities</u> , such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c Lifting or carrying groceries	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d Climbing several flights of stairs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e Climbing one flight of stairs.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f Bending, kneeling, or stooping.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g Walking <u>more than a mile</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h Walking <u>several blocks</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i Walking <u>one block</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j Bathing or dressing yourself.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	Yes ▼	No ▼
a Cut down on the amount of time you spent on work or other activities	<input type="checkbox"/>	<input type="checkbox"/>
b <u>Accomplished</u> less than you would like.....	<input type="checkbox"/>	<input type="checkbox"/>
c Were limited in the kind of work or other activities	<input type="checkbox"/>	<input type="checkbox"/>
d Had difficulty performing the work or other activities (for example, it took extra effort)	<input type="checkbox"/>	<input type="checkbox"/>

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	Yes ▼	No ▼
a Cut down on the amount of time you spent on work or other activities	<input type="checkbox"/>	<input type="checkbox"/>
b <u>Accomplished</u> less than you would like.....	<input type="checkbox"/>	<input type="checkbox"/>
c Did work or other activities less carefully than usual.....	<input type="checkbox"/>	<input type="checkbox"/>

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

Not at all ▼	Slightly ▼	Moderately ▼	Quite a bit ▼	Extremely ▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. How much bodily pain have you had during the past 4 weeks?

None ▼	Very mild ▼	Mild ▼	Moderate ▼	Severe ▼	Very Severe ▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

Not at all	A little bit	Moderately	Quite a bit	Extremely
▼	▼	▼	▼	▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...

	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
	▼	▼	▼	▼	▼	▼
a Did you feel full of pep?.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b Have you been a very nervous person?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c Have you felt so down in the dumps that nothing could cheer you up?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d Have you felt calm and peaceful?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e Did you have a lot of energy?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f Have you felt downhearted and blue?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g Did you feel worn out?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h Have you been a happy person?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i Did you feel tired?.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
▼	▼	▼	▼	▼
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. How TRUE or FALSE is each of the following statements for you?

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false
	▼	▼	▼	▼	▼
a I seem to get sick a little easier than other people.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b I am as healthy as anybody I know.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c I expect my health to get worse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d My health is excellent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Thank you for completing these questions!

SF-36® Health Survey 1988, 2002 by Medical Outcomes Trust and QualityMetric Incorporated. All Rights Reserved.

SF-36® is a registered trademark of Medical Outcomes Trust.
(SF-36 Standard, US Version 1.0)

Source: <http://www.qualitymetric.com/products/SurveyInfo.shtml>

References

- Appleby J (1996). 'Promoting efficiency in the NHS: problems with the labour productivity index.' *BMJ*, vol 313, pp 1319–21.
- Appleby J, Boyle S, Devlin N, Harley M, Harrison A, Locock L, Thorlby R (2004). *Sustaining Reductions in Waiting Times: Identifying successful strategies*. Final report to the Department of Health. London: King's Fund (in press).
- Appleby J, Harrison A, Devlin NJ (2003). *What is the Real Cost of More Patient Choice?* London: King's Fund.
- Appleby J, Little V (1993). 'Health and efficiency.' *Health Services Journal*, vol 103, pp 20–2.
- Appleby J, Sheldon T, Clarke A (1993). 'Run for your money.' *Health Services Journal*, vol 103, pp 22–4.
- Bellamy N, Buchanan WW, Goldsmith CH, Stitt LW (1988). 'Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip or knee.' *Journal of Rheumatology*, vol 15, pp 1833–40.
- Bowling A (2001). *Measuring Disease*, 2nd ed. Buckingham: Open University Press.
- Bowling A (1997). *Measuring Health*, 2nd ed. Buckingham: Open University Press.
- Brazier J, Akehurst R, Brennan A, Dolan P, Klaxton C, McCabe C, O'Hagan T, Sculpher M, Tsuchiya A (2004). *Should patients have a greater role in valuing health states: whose well-being is it anyway?* Paper presented to the joint CES/HESG conference, Paris, January 2004.
- Buchner D, Graboys TB, Johnson K, Mordin MM, Goodman L, Partsch DS, Goss TF (2001). 'Development and validation of the ITG health-related quality of life of short-form measure for use in patients with coronary artery disease.' *Clinical Cardiology*, vol 24, pp 556–63.
- Burge P, Devlin NJ, Appleby J, Rohr C, Grant J (2004). *Do patients always prefer shorter waiting times? A discrete choice analysis of patients' stated preferences in the London Patient Choice Project (LPCP)*. Paper to be presented to the 2004 European Health Economics Conference, LSE.
- BUPA (2003). BUPA's clinical outcomes programme. Internal document, March 2003.
- CHI (2002). *NHS Star Ratings* (accessible at: <http://www.chi.nhs.uk/ratings/>)
- Clarke A, McKee M, Appleby J, Sheldon T (1993). 'Efficient purchasing.' *BMJ*, vol 307, pp 1436–7.
- Crosby RD, Kolotkin RL, Williams GR (2003). 'Defining clinically meaningful change in health-related quality of life.' *Journal of Clinical Epidemiology*, vol 56, pp 395–407.
- Dennison CR (2002). 'The role of patient-reported outcomes in evaluating the quality of oncology care.' *American Journal of Managed Care*, pp S580–6.
- Department of Health (2000) High Level Performance Indicators. (accessible at: <http://www.performance.doh.gov.uk/nhsperformanceindicators/indicators2000.htm>)
- Department of Health (2001). *The Expert Patient: A new approach to chronic disease management for the 21st century*. London: Department of Health.
- Department of Health (2003). *Building on the Best: Choice, responsiveness and equity in the NHS*. London: Department of Health.

- Derrett S, Devlin NJ, Hansen P, Herbison P (2003). 'Prioritising patients for elective surgery: a prospective study of clinical priority assessment criteria in New Zealand.' *International Journal of Technology Assessment in Health Care*, vol 19 (1), pp 91–105.
- Derrett S, Devlin N, Harrison A (2002). 'Waiting in the NHS Part 2: a change of prescription.' *Journal of the Royal Society of Medicine*, vol 95, pp 280–3.
- Derrett S, Paul C, Herbison P, Williams H (2002). 'Evaluation of explicit prioritisation for surgery: a prospective study.' *Journal of Health Services Research and Policy*, vol 7, suppl 1, pp S14–22.
- Desai P, Reidy A, Minassian DC, Vafidis G, Bolger J (1996). 'Gains from cataract surgery: visual function and quality of life.' *British Journal of Ophthalmology*, vol 80, pp 868–73.
- Devlin NJ, Appleby J, Parkin D (2003). 'Patients' views of explicit rationing: what are the implications for health service decision-making?' *Journal of Health Services Research and Policy*, vol 8 (3), pp 183–6.
- Dodrill CB (1978). 'A neuropsychological battery for epilepsy.' *Epilepsia*, vol 19, pp 611–23.
- Dodrill CB, Batzel LW, Quiesser HR, Temkin NR (1980). 'An objective method for the assessment of psychological and social problems among epileptics.' *Epilepsy*, vol 21, pp 123–35.
- Dr Foster (2004) *Hospital Guides* (accessible at: <http://www.drfooster.co.uk/>)
- Drummond MF, O'Brien B, Stoddart G (1997). *Methods for the Economic Evaluation of Health Care Programmes*, 2nd ed. Oxford: Oxford Medical Publications.
- Endicott J, Spitzer RL (1978). 'A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia.' *Archives of General Psychiatry*, vol 35, pp 837–44.
- Essink-Bot ML, Krabbe PFM, Bonsel GJ, Aaronson NK (1997). 'An empirical comparison of four generic health status measures: The Nottingham Health Profile, the Medical Outcomes Study 36 Item Short Form, the COOP/WONCA Charts, and the EuroQol Instrument.' *Medical Care*, vol 35, pp 522, 537.
- Garratt AM, Ruta D (1996). 'Measuring health outcomes in low back pain' in *Lumbar Spine Disorders, Current Concepts*, Aspen RM and Porter RW eds. Singapore: World Scientific.
- Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R (2002). 'Quality of life measurement: bibliographic study of patient assessed health outcome measures.' *BMJ*, vol 324, pp 1417–19.
- General Medical Council (2001) *Good Medical Practice*. www.gmc-uk.org
- Grundy SM (2002). 'Statin therapy in older persons.' *Archives of Internal Medicine*, vol 162, pp 1239–1331.
- Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. (1989). 'Development and testing of a new measure of health status for clinical trials in heart failure.' *Journal of General Internal Medicine*, vol 4(2), pp 101–7.
- Hadorn DC, Holmes AC (1997). 'The New Zealand priority criteria project. Part 1: Overview.' *BMJ*, vol 314, pp 131–4.
- Harrison A (2003). *Getting the Right Medicines? Putting public interests at the heart of health-related research*. London: King's Fund.
- Harrison A, New B (2002). *Public Interests, Private Decisions: Health-related research in the UK*. London: King's Fund.

Healthcare Commission (2004a). 2004 *Performance Indicators*.
http://ratings.healthcarecommission.org.uk/indicators_2004/

Healthcare Commission (2004b). *Investigation reports published by the Healthcare Commission* (accessible at:
http://www.healthcarecommission.org.uk/YourLocalHealthServices/NHSTypeOfReport/fs/en?CONTENT_ID=4002761&chk=lfkAdn)

Her Majesty's Treasury (2002a). *Public Service Agreements 2002–2006*. London: The Stationery Office. http://www.hm-treasury.gov.uk/Spending_Review/spend_sr02/psa/spend_sr02_psaindex.cfm

Her Majesty's Treasury (2002b). *Department of Health Value for Money Target*. http://www.hm-treasury.gov.uk/performance/targets/perf_target_21.cfm

Hibbard JH (2003). 'Engaging healthcare consumers to improve the quality of care.' *Medical Care*, vol 41 (1), S61–S70.

House of Commons Health Committee (1997). *The specific health needs of children and young people*. London: The Stationary Office.

Ireson CI, Ford MA, Hower JM, Schwartz RW (2002). 'Outcome report cards: a necessity in the health care market.' *Archives of Surgery*, vol 137, pp 46–51.

Juniper EF, Buist AS, Cox FM, Cox FM, Ferrie PJ, King DR (1999). 'Validation of a standardised version of the Asthma Quality of Life Questionnaire.' *Chest*, vol 115, pp 1256–70.

Kind P, Williams AH (2004). 'Measuring success in healthcare: the time has come to do it properly!' *Health Matters*, issue 9, University of York.
<http://www.york.ac.uk/healthsciences/pubs/hpmindex.htm>

Lamping DL, Rowe P, Clarke A, Black N, Lessof L (1998). 'Development and validation of the menorrhagia outcomes questionnaire'.
British Journal of Obstetrics and Gynaecology, vol 105, pp 766–79.

Leatherman S, Hibbard JH, McGlynn EA (2003). 'A research agenda to advance quality management and improvement'. *Medical Care*, vol 41 (1), pp S80–S86.

London Modernisation Board (2004). *Modernising London's NHS: London Patients Choice Project: Project overview*. London: London Modernisation Board.
<http://www.london.nhs.uk/patientchoice/> (accessed 28 June 2004)

Mannion R, Goddard M (2001). 'Impact of published clinical outcomes data: case study in NHS hospital trusts'. *BMJ*, vol 323, pp 260–3.

Mant J, Hicks N (1995). 'Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating myocardial infarction'.
BMJ, vol 311, pp 793–6.

Marshall MN, Shekelle PG, Leatherman S, Brook RH (2000). 'The public release of performance data: what do we expect to gain? A review of the evidence.'
JAMA, vol 283 (14), pp 1866–74.

Maynard A, Bloor K (2003). 'Do those who pay the piper call the tune?' *Health Policy Matters*, issue 8, October 2003. www.york.ac.uk/depts/hstd/pubs/hpmindex.htm

Meenan RF, Gertman PM, Mason JH (1980). 'Measuring health status in arthritis: the arthritis impact measurement scales.' *Arthritis and Rheumatism*, vol 23 (2), pp 146–52.

Meenan RF, Mason JH (1990). *AIMS2 Users' Guide*. Boston MA: Boston University School of Medicine, Department of Public Health. 0

- National Centre for Health Outcomes Development (2000). *Outcome Indicators for Cataract*. Report of a working group to the Department of Health. Oxford: PHI database, Oxford University.
- NICE (2004). *Guide to the Methods of Technology Appraisal*. London: NICE.
- Nightingale F (1863). *Notes on hospitals*, 3rd ed. London: Longmans.
- Radloff LS (1977). 'The CES-D scale: a self-report depression scale for research in the general population.' *Applied Psychological Measurement*, vol 1, pp385–401.
- Royal College of Physicians (2004). *Consultant-level comparisons*. Health Informatics Unit. www.rcplondon.ac.uk/college/hiu
- Ruta DA Garratt AM, Wardlaw D, Russell IT (1994). 'Developing a valid and reliable measure of health outcomes for patients with low back pain.' *Spine*, vol 19, pp 1887–96.
- Schag CAC, Heinrich RL, Ganz PA (1983). 'Cancer Inventory of Problem Situations: an instrument for assessing cancer patients' rehabilitation needs'. *Journal of Psychosocial Oncology*, vol 1, pp 11–24.
- Secretary of State for Health (2000). *The NHS Plan: A plan for investment, a plan for reform*. Cm 4818-I. London: The Stationery Office.
- Smee C (1991). 'How might measures of quality of life be useful to me as a health economist?' in *Measures of Quality of Life and the Uses to which such Measures may be Put*, Hopkins A ed. London: Royal College of Physicians.
- Stead ML, Crocombe WD, Fallowfield LJ, Selby P, Perren TJ, Garry R, Brown JM (1999). 'Sexual activity questionnaires in clinical trials: acceptability to patients with gynaecological disorders.' *British Journal of Obstetrics and Gynaecology*, vol 106, pp 50–4.
- Steinberg EP, Tielsch JM, Schein OD, Javitt JC, Sharkey P, Cassard SD, Legro MW, Diener WM, Bass EB, Damiano AM, Steinwachs DM, Sommer A (1994). 'The VF-14: an index of functional impairment in patients with cataract.' *Archives of Ophthalmology*, vol 112, pp 630–8.
- Stevens S (2004). 'Reform strategies for the English NHS.' *Health Affairs*, vol 23 (3), pp 37–44.
- Treasury Committee (2003). *Third Report: The 2003 pre-budget report. 2003*. <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmtreasy/136/13602.htm>

