



# Measure for measure

Using outcome measures  
to raise standards in the NHS

Richard Hamblin and Janan Ganesh  
Edited by Gavin Lockhart

policy  
exchange

Think Tank of the Year 2006/2007

---

# Measure for measure

Using outcome measures to raise standards in the NHS

Richard Hamblin and Janan Ganesh

Edited by Gavin Lockhart



Policy Exchange is an independent think tank whose mission is to develop and promote new policy ideas which will foster a free society based on strong communities, personal freedom, limited government, national self-confidence and an enterprise culture. Registered charity no: 1096300.

Policy Exchange is committed to an evidence-based approach to policy development. We work in partnership with academics and other experts and commission major studies involving thorough empirical research of alternative policy outcomes. We believe that the policy experience of other countries offers important lessons for government in the UK. We also believe that government has much to learn from business and the voluntary sector.

#### **Trustees**

Charles Moore (Chairman of the Board), Theodore Agnew, Richard Briance, Camilla Cavendish, Richard Ehrman, Robin Edwards, George Robinson, Tim Steel, Alice Thomson, Rachel Whetstone.

---

---

## About the authors

### **Richard Hamblin**

*Harkness Fellow, Group Health Co-operative of Puget Sound*

Richard is a 2006-07 Commonwealth Fund Harkness Fellow in Healthcare Policy and Practice, based in Seattle at the Centre for Health Studies. His work there concentrates on the way that publicising information about the quality of health services may improve them. Previously he was head of analytic support in the Healthcare Commission, the English health services regulator. His primary responsibility there was ensuring a rigorous analytic and evidential base for regulatory activities. Richard's previous career encompasses NHS management and health services research, including a spell as a research officer at the King's Fund where he produced studies on London's healthcare system, the effect of the introduction of the internal market into the NHS in the 1990s and methods of modelling waiting-list dynamics. His particular interest is in health informatics and the use of rigorous analysis of performance data as a tool for improvements in quality and efficiency of health services. He is a graduate of

University College London (1992). He writes here in a personal capacity and views expressed are not necessarily those of the Commonwealth Fund, its directors, officers or staff.

### **Gavin Lockhart**

*Research Director, Policy Exchange*

Gavin is Research Director at Policy Exchange with responsibility for healthcare and police research. He studied Sociology at Edinburgh University. After graduating in 2002 with a first-class degree, Gavin worked as a management consultant before joining Policy Exchange in August 2006.

### **Janan Ganesh**

*Research Fellow, Policy Exchange*

Janan is a Researcher at Policy Exchange. He was born in 1982 and graduated top of his year from Warwick University (BA Politics) before studying public policy at UCL, where he focused on health. In 2006, he co-authored *Compassionate Conservatism* for Policy Exchange and won the T.E. Utley Memorial Prize for young journalists. He has also worked for Zac Goldsmith.

---

© Policy Exchange 2007

Published by

Policy Exchange, Clutha House, 10 Storey's Gate, London SW1P 3AY

[www.policyexchange.org.uk](http://www.policyexchange.org.uk)

ISBN: 978-1-906097-07-3

Printed by Heron, Dawson and Sawyer

Designed by JohnSchwartz, [john@thefrontline.net](mailto:john@thefrontline.net)

---

## Contents

	Executive Summary	5
1	Introduction	9
2	The current situation	13
3	From information to action: five case studies	20
	<i>Cardiac Surgery Reporting System in New York State</i>	20
	<i>Clinical Resource and Audit Group Outcome Indicators Reports</i>	23
	<i>Veterans Health Administration National Surgical Quality Improvement Program</i>	25
	<i>Healthcare Commission Star Ratings</i>	30
	<i>The London Patient Choice Project</i>	32
4	Patients, payment and providers	35
5	Technical difficulties	42
6	Patient Reported Outcome Measures (PROMs)	47
7	Outcome Measurement in Primary Care	51
8	Recommendations	57
	Appendix one: explaining why US comparisons are valid	67
	Appendix two: information for accountability	69
	Appendix three: information for patient choice and activation	71
	Glossary	72

---

## Acknowledgements

Policy Exchange would like to thank John Nash, Hugh Osmond, Henry Pitman and Mike Sherwood who supported this research project. We would also like to thank the other donors who generously and anonymously supported the project.

Special thanks go to Professor Gwyn Bevan, Mohammed Mohammed, Philippa Ingram and all those who commented on early versions of this report. The Policy Exchange Health and Social Care Steering Group also provided invaluable guidance.

The views expressed are not necessarily those of members of the Policy Exchange Health and Social Care Steering Group or of those who commented on drafts of the report.

---

## Executive Summary

### From information to action: context

The last two decades have been characterised by a revolution in the availability of information. This has included a transformation of the capacity to store, retrieve and disseminate data (including, but not limited to, the growth of the internet), the questioning of previously accepted intellectual authority,<sup>1</sup> and more coherent theorising about the value of information to society and the uses of information as a resource.<sup>2</sup> The effects of this revolution are as noticeable in healthcare as anywhere else.

There are, broadly, two different types of data about health and healthcare that are increasingly available to the public. The first is about diseases and remedies,<sup>3</sup> and the second is about the quality and efficiency of health services. Although clearly different in focus, both assume a view of patients that transcends the traditional perception of them as passive recipients of care and information, seeing them instead as more pro-active consumers of services, motivated to learn more about their illnesses, their treatments and most importantly about the quality of the carers and care providers.

At its foundation, the NHS agreed a 'bargain' with doctors in which they accepted budgetary constraints in return for clinical autonomy and freedom from outside scrutiny. The key to this deal was its assumption of altruism and professionalism on the part of clinicians. Until the late 1980s, the image of the NHS was that its resources were limited and directed at ensuring good clinical care, and that its outcomes were as good as, if not better than, those achieved in comparable countries.

This impression was shattered in the 1990s by evidence from international comparisons which showed that the UK had poor outcomes for cancer, heart dis-

ease and stroke – the country's principal burden of disease. In addition, a series of scandals involving incompetent (and, in some cases, malign) doctors attracted dreadful publicity. Particularly shocking was that these abuses were allowed to continue unchecked for long periods.

“ The Government and Department of Health currently focus on the ‘wrong’ measurements of performance ”

The Government's response was to develop policies that put quality at the heart of the programme of NHS reform. Using information has been an essential part of this strategy. For example, a range of national clinical audits was introduced to strengthen clinical governance, a Modernisation Agency was created and the NHS Information Centre was established. The NHS has a huge amount of data. But the Government and Department of Health currently focus on the 'wrong' measurements of performance.

Modern enterprises see information on their performance as vital – what is extraordinary about the NHS is that it spends so much on hospital care and knows so little about what this spending achieves. Outcome data currently focuses on mortality and readmission. This excludes around 90 per cent of hospital admissions. Whenever information has been collected on healthcare, it has revealed serious failings that require corrective action, and has identified high performers from whom others can learn.

*Measure for measure* focuses on information about the quality of health services and explores ways in which it can lead to better quality healthcare for all. We argue that the information we need is not usually available and that, even if it were, it

1. Giles J, "Internet encyclopaedias go head to head", *Nature*, 2005

2. Such as Cleveland H, "Information as Resource", *The Futurist*, December 1982

3. Sometimes described as 'information therapy' this is increasingly available through various media (<http://www.information-therapy.org/>) and has led to the concept of the 'expert patient'

would not by itself lead to the improvement of health services. If information alone were enough, none of the scandals identified below would have been allowed to go on for so long.

In chapter two we explain that there are different ways in which the publication of data may lead to quality improvement. One is that providers themselves actively respond, encouraged by incentives to improve their performance.<sup>4</sup> Another is that the changed behaviour of patients can drive up standards. Patients may choose better providers, but may also become more confident in seeking improved services from their existing provider. But using information to measure quality is problematic. It is more difficult to assess the quality of primary care<sup>5</sup> than secondary care.<sup>6</sup> Chronic care, which accounts for much of primary care, often involves multiple agencies (not just the GP), multiple interventions (not just a one-off operation) and, by definition, sees declining health outcomes for patients over time. There is an understandable enthusiasm for the measurement of primary and chronic care to be developed further, and the quality and outcomes framework (QOF) is a strong foundation on which to build. This does not mean, however, that the issue of hospital measurement has been resolved.

There remain obstacles to using information in helpful ways. One of these is getting the level of information correct. The practice of giving overall ratings for individual hospital trusts, followed by the Government since 2000, has led to simplistic assessments and has left gaps in the information available below the level of the hospital. For example, to know that a hospital has low mortality rates for medical and surgical admissions tells us nothing about sub-specialties within surgery.<sup>7</sup>

Given the present moves towards transparency and external quality measurement, the publication of data is no longer contro-

versial. Indeed, it has come to be seen as inevitable. The question remains, however: how should it be done to ensure the maximum benefit for patients?

### Learning from previous measurement schemes

There have been a number of measurement schemes in the UK and US that provide useful lessons for the future. In chapters three and four we consider case studies of attempts to use information to improve quality (three from the United States, two from England and one from Scotland) and conclude as follows:

- to be successful, the measures themselves must be both visible and credible for clinicians; those that are not will be ignored;
- information that is primarily used internally can be used to improve care (as in the case of the Veterans' Administration in the US) when there is a robust improvement programme that includes on-the-ground investigation to identify and correct problems, and sharing of good practice;
- publishing the measures, rather than just reporting confidentially back to providers, seems to increase the likelihood of changing their behaviour, as was noticed in the QualityCounts initiative in Wisconsin US. Although, as the New York case study shows, there is increased concern about gaming when measures are made public;
- the effects of linking formal incentives to quality measurement, such as pay for performance, are ambiguous. Incentives encourage unintended perverse responses such as gaming and falsification of data. It is not clear that the extra costs of providing a financial incentive encourage sufficiently better performance;
- research, including our own survey, shows that patients want quality of care

4. Formal incentives here would be payment mechanisms or other sanctions and rewards; informal incentives would be kudos and censure for individuals and organisations.

5. The frontline of the NHS is officially called primary care. The initial contact for many people when they develop a health problem is with a member of the primary care team, usually their GP. Many other health professionals work as part of this frontline team: nurses, health visitors, dentists, opticians, pharmacists and a range of specialist therapists. NHS Direct and NHS walk-in-centres are also primary care services.

6. Specialised medical services and commonplace hospital care (outpatient and inpatient services). Access is often via referral from primary healthcare services.

7. There are also organisational issues that need to be addressed. For example, the primary role of the Healthcare Commission is to assess how well NHS hospitals perform against Government targets and standards, and not to provide information below that level, while lower level assessments of individual services are available, for example information about diabetes services has recently been released. And despite the emergence of other providers of information, gaps in data remain.

information, yet it is not clear that patient choice is necessarily stimulated by publication of information. Very well studied publication schemes in the US, such as QualityCounts and cardiac surgery mortality in New York State showed that publishing data did not lead to changes in hospital market share. The London Patient Choice Project, by contrast, appears to have been successful in encouraging choice, although this may reflect both the large infrastructure put in place to support choice, and the relatively simple metrics needed to understand how long one waits (a somewhat easier proposition than comparing the quality of two services).

#### Technical issues

Given these examples of apparently successful use of measurement, one may wonder why the practice is not more widespread. Part of the answer is that it is hard to put into action. There are difficulties in establishing measurement regimes that stimulate improvements without also producing unintended consequences. These include the design of the incentive system around the measurement system; ensuring adequate data quality; and designing analysis that is sufficiently sophisticated to allow for the complexities of healthcare, and is easily understood by its target audience. There are also problems in deciding where and how to publish the results. These difficulties are technical rather than ideological, and are discussed in chapter five.

#### Lessons for reform

If we are serious about improving the quality of care, then we need much more precise information than the limited amount that is available at present. Nothing is known about the outcomes of most patients discharged from NHS hospitals. It

has been argued that patient-reported measures offer an important adjunct to clinicians in the care of their patients.<sup>8</sup> Self-completed questionnaires with adequate measurement properties offer a quick way for patients to provide evidence of how they view their health – evidence that can complement existing clinical data. Instruments applied in this context can be used to screen for health problems and to monitor progress of health problems identified, as well as the outcomes of any treatment.

“ Patient-based outcome measures may also help change the culture of the NHS; an organisation which is far from universally patient-focused. It is a politically-led organisation, where the Government, not the patient, is the paymaster ”

As we argue in chapter six, patient-based outcome measures may also help change the culture of the NHS; an organisation which is far from universally patient-focused. It is a politically-led organisation, where the Government, not the patient, is the paymaster.

At the end of this report we present a vision of the potential future landscape of healthcare quality information that reforms could facilitate and we make eighteen recommendations. We believe that such a landscape would provide accountability, promote increased patient trust, and improve performance.

The information needed falls into three categories:

#### *Information for accountability*

Accountability measures are designed to assure taxpayers that their local health services (including NHS providers, Independent Treatment Centres and private providers doing publicly funded work) do two things:

9. Tarlov AR et al, "The Medical Outcomes Study: An application of methods for monitoring the results of medical care", *Journal of the American Medical Association*, 1989; Nelson EC, Berwick DM, "The measurement of health status in clinical practice", *Medical Care*, 1989



- provide at least a minimum acceptable level of care, including safety, access, clinical competence, and compassion;
- spend public money with due care and consideration and achieve the goals which the Government decide are appropriate.

*Information for patient choice and activation*

Information is often seen as facilitating patient choice, but it also serves activation, the patients' willingness to be involved and assertive in the decision-making in their own care. Moreover, choice can only really apply to predictable elective care, and especially surgery. It is not generally relevant to emergency care or to management

of long-term conditions, where it will tend to conflict with continuity of care. Whether for choice or activation, releasing small amounts of information about the quality of specific services to patients with specific conditions, rather than consumers generally, will provide invaluable support.

*Information for providers*

This information is used to activate providers' intrinsic motivations of professionalism and altruism. It does not need to be published but does require a mechanism to share the information with providers and hospitals, and to encourage and monitor improvement.

---

# 1

## Introduction

This chapter sets out the recent historical context of reform in the NHS, the different ways in which information can be used to drive quality improvement and the information that is currently available.

### Context: information and governance

At the start of the 1990s the NHS was still recognisable as the organisation that was created after the Second World War. This was despite the introduction of general managers in the 1980s and an internal market from 1991. As an institution, it enjoyed public esteem; doctors were trusted by their patients to do the best they could within the resources that were available. The image of the NHS was one whose limited resources were directed at ensuring good clinical care, and whose outcomes were as good as, if not better than, those achieved in comparable countries. This image was shattered in the 1990s, as evidence emerged of relatively poor outcomes for cancer, heart disease and strokes in the England and Wales – the principal causes of the country’s burden of disease.

In addition, a series of scandals involving patients suffering at the hands of incompetent (and, in some cases, malign) doctors attracted dreadful publicity. Particularly shocking was that these abuses were allowed to continue unchecked for so long. One of the most notorious cases was at the Bristol Royal Infirmary, where an exceptionally high number of babies died after heart surgery. The ‘Kennedy Report’ of the public inquiry into the Bristol case identified the root cause of such inaction as the absence of any system to govern and regulate quality of care in the

NHS.<sup>9</sup> The report observed that individuals who were aware of the problem did not believe it was their duty to act.

We give evidence in this report of serious lapses in the quality of care in the NHS that came to light in the 1990s and consider the policies that have been introduced over the past ten years to remedy them. Before considering what we can learn from experience in the UK and elsewhere, it is necessary to understand why the NHS did so little to analyse, monitor and govern the quality of its care.

Like many public services in North America and Western Europe, the NHS has traditionally been a “network” model of public service delivery. This is a system where the State funds and provides a service, and trusts frontline workers to perform their roles with altruism and professional competence.<sup>10</sup>

The lesson much of the British elite and electorate took from the nation’s successful struggle against Nazi Germany was that the liberal individualism of the Victorian era should be replaced by the kind of high-minded collectivism that won the war. This collectivism, embodied by the Beveridge Report of 1942, was reinforced by the experience of the emergency medical service during the war: many from voluntary hospitals worked outside London and were shocked by what they found.

This faith in public sector workers led to the notion that scrutiny of their performance and incentives for improvement were unnecessary. It was enough to rely on the public service ethos of those who staffed the welfare state to provide the best possible per-

9. Learning from Bristol - Report of the Public Inquiry into Children’s Heart Surgery at the Bristol Royal Infirmary [Kennedy Report], the Stationery Office, 2001

10. Le Grand J, Motivation, Agency and Public Policy: Of knights and knaves, pawns and queens, Oxford University Press, 2003

formance. And nowhere did this assumption run deeper than in the NHS.

The “deal” between NHS workers and the Government whereby the latter grants the former a greater or lesser degree of independence in discharging their duties is an example of what Christopher Hood has called “public sector bargains”. These are “explicit or implicit agreements between public servants – the civil or uniformed services of the state – and those they serve. The other partners in such bargains consist of politicians, political parties, clients, and the public at large.”<sup>11</sup> Public sector bargains are part of the “living constitution” of a state. As Hood remarks: “Any meaningful democracy can exist only if at least the police and armed forces show some allegiance to the elected Government”.<sup>12</sup> But these bargains extend to others professions as well, and include health workers.

Hood distinguishes between two broad categories of public sector bargain. One is the “principal-agent bargain”, in which public servants are seen as rational actors who are willing to subvert or ignore the interests of the Government or the general public if their own interests so require. The NHS, however, was organised on the basis of the second type, the “trustee-type bargain”. This is derived from the legal concept of property being put under the charge of a “trust” for the benefit of others. Trustees are not controlled by their beneficiaries, whose powers are limited to being able to take legal action to ensure that the trustees comply with the terms of the trust. A trustee bargain, therefore, is one in which public servants (the trustees) are afforded wide latitude to determine what is in the interests of the general public (the beneficiaries).

Under such a bargain, “the tenure and rewards of public servants are not under the direct control of those for whom they act, the skills and competencies they are expected to show are not determined by the instrumental interests of elected politicians, and loyalty lies to an entity that is broader than

the elected Government of the day.”<sup>13</sup> The NHS has many of the characteristics of a trustee-type bargain.<sup>14</sup>

Rudolf Klein has suggested that, from the start of the NHS, there was an implicit concordat between the state and the medical profession, in which clinicians accepted budgetary constraint in return for clinical autonomy.<sup>15</sup> In Julian Le Grand’s words, “politicians and civil servants (allocated) resources at macro level and gave medical professionals almost complete clinical freedom to make ground-level decisions as to which patients should receive what treatment”.<sup>16</sup> They were, to employ Le Grand’s famous distinction, “knights” rather than “knaves”. A corollary was that it was unnecessary to scrutinise the performance of those who worked in public services. The caring professions were assumed to be a model of “knightly” behaviour.

So, from its creation in 1948, the NHS has essentially remained a network organisation, with clinical autonomy over care integral to its organisational DNA. There are many advantages to this, not least high morale for professionals and low administrative costs.

However, there are two bold assumptions behind such an approach to health-care: that doctors are all “knights” who put their patients’ interests before their own and that they all also routinely perform to acceptable levels of competence. The experiences of the 1990s challenged this assumption profoundly.

### The crisis in quality

Whatever the organisational changes that took place over the first 50 years of the NHS, there was no challenge to the theory of professional self-regulation and the exclusion of outside scrutiny of quality of care. It is important to recognise that this organisational pattern, reflecting as it does the trust of the public in the caring professions, is the

11. Hood C and Lodge M, *The Politics of Public Service Bargains: Reward, Competency, Loyalty - and Blame*, Oxford University Press, 2006

12. *Ibid*

13. *Ibid*

14. Such as the broad discretion afforded to clinicians, in the absence of systematic scrutiny and monitoring of performance until relatively recently, and in its ongoing failure to attempt to measure the extent to which the health of patients improves as a result of treatment.

15. Klein R, *The New Politics of the National Health Service*, Radcliffe Press: Oxford, 2006

16. Le Grand, 2003

explanation for the crisis of quality that threatened the NHS in the 1990s.

This pattern was so deeply rooted in the NHS that the three fundamental reorganisations in the 1970s, 1980s and 1990s simply entrenched it further. The first significant reorganisation of the NHS, in 1974,<sup>17</sup> was designed on the principle of consensus management, which held that no decision could be made without the unanimous agreement of all members of the management teams. This meant that medical representatives could veto change. Mrs Thatcher's Government sought to end consensus management in the 1980s with the introduction

of general managers.<sup>18</sup> However, these managers had no responsibility for the quality of care, which continued to be seen as a professional prerogative. Then, alongside the creation of the internal market from 1991, the Thatcher Government introduced medical audit. However, medical audit proved singularly ineffective in the Bristol case, where the 'Kennedy Report' found it was outside the normal managerial processes and lacked a presence in the area of children's heart surgery.

Table 1 gives five examples of scandals that came to light in the 1990s and that were the subject of inquiries commissioned by the

17. Management Arrangements for the Reorganised Health Service, Department of Health and Social Security, HMSO, 1972

18. Pollitt C, Harrison S, Hunter D, Marnoch G, "General Management in the NHS: The Initial Impact 1983-88", Public Administration, 1991

**Table 1: delays in responding to a series of scandals**

Scandal	Nature	First concerns raised	Action taken
Failures of children's heart surgery at the Bristol Royal Infirmary	1991-1995 at least 30 more children died than would have been expected in a typical unit	1986 by South Glamorgan Health Authority	1995 DoH advises that an outside independent inquiry is essential; 1997 inquiry reports
Actions of gynaecologist Rodney Ledward	Found guilty of bungling 13 operations and allegedly damaging hundreds more women	1986 senior management is made aware of Mr Ledward's complication rate and his cavalier manner	1996 suspended and later dismissed by South Kent Hospitals NHS Trust; 1998 struck off medical register; 1999 Ritchie inquiry set up 2000 Ritchie inquiry reports
Organ-stripping at Alder Hey Children's Hospital, Liverpool, by senior pathologist	Large number of hearts from deceased children retained without the consent of their parents	1989 onwards, Alder Hey and the university missed numerous opportunities to discipline Professor Dick van Velzen	1999 Redfern inquiry announced 2001 Redfern inquiry reports
Yorkshire GP, Harold Shipman, administered lethal drugs to his patients	1975-1998 he may have killed more than 200 patients	1998 a GP in a neighbouring practice reported her concerns to the local Coroner	1998 charged with murder following forgery of a will 2000 convicted of 15 counts of murder
Indecent assaults by Loughborough GP Peter Green	1980s-90s described as preying on young male patients for 17 years	1985 patients tried to raise concerns with the family health service authority	2000 jailed for eight years for nine counts of indecent assault

Secretary of State for Health in the period 2000–2002. In each case, individual doctors abused the trust that patients had placed in them, and in each case there was an extraordinary reluctance from other parties to act on the information supplied. In the Shipman case, the failure of the police inquiry to act on suspicion of murder was fortunately followed soon after by the arrest of Shipman, who drew attention to himself by incompetently forging a will.<sup>19</sup> But that failure illustrates the general problem of disbelief in what the evidence suggested to be the case: that a popular GP was murdering his patients. In each case medical behaviour was an anomaly at variance with the assumption that it is right to trust all doctors to act in the best interests of their patients.

In addition to these scandals over lapses by a few individuals, there was evidence from international comparisons of two different kinds of systemic failing. First, the UK population had relatively poor outcomes for cancers, heart disease and strokes. Secondly, they waited much longer for hospital care than patients in other countries with national health services.

The Labour Government's response to these twin failings was to develop policies that put quality at the heart of its programme of NHS reforms. In 2003, two experts characterised the quality agenda for the NHS in England as being the "most ambitious, comprehensive, systemic and intentionally funded effort to create predictable and sustainable capacity for improving the quality of a nation's health-care system".<sup>20</sup> Using information has been an essential part of this strategy. Developments have included a range of national clinical audits to strengthen clinical governance, as well as the creation of the Modernisation Agency and the NHS Information Centre.<sup>21</sup>

Over the past decade and a half, the assumption that doctors will inevitably be altruistic and competent has been disproved by events. As a result the principle of external measurement of quality in UK health services is no longer controversial, and to many seems inevitable. But the question remains: how should quality measurement be carried out to ensure the maximum benefit for patients?

19. The Shipman Inquiry, First Report: Death Disguised, Secretary of State for Health, 2002

20. Leatherman S, Sutherland K, The Quest for Quality in the NHS: A Mid-Term Evaluation of the Ten-Year Quality Agenda, the Stationery Office, 2003 and Leatherman S, Sutherland K, "Quality of care in the NHS of England", 2004

21. <http://www.ic.nhs.uk/> (Last accessed 03.05.07)

---

# 2

## The current situation

### Summary

This section of the report describes what information we currently have, what information we need, what effect the publishing of information might have, and the obstacles that must be overcome if information is to drive greater accountability and quality.

### How information can be used to improve quality of care

There are two reasons for measuring performance and putting the resulting information in the public domain. The first is to increase the accountability of healthcare organisations and professionals. The second is to stimulate improvements in the quality of care provided.<sup>22</sup> The fact that publication leads to increased accountability is self-evident yet the way in which improvement occurs is complex and is explored below.

There are at least four different mechanisms whereby making information public may effect improvement:

- Information may help patients to “shop around” for the best provider or institution;
- Patients may use the information to give themselves greater confidence and power in negotiating the health-care system without necessarily changing the provider or institution they go to;
- Purchasers of healthcare may make better and more targeted decisions about where to purchase care;

- Providers and institutions may themselves be motivated to improve regardless of any other incentive, by virtue of the kudos that a good report imparts.

“ Patients may use the information to give themselves greater confidence and power in negotiating the health-care system”

However, whatever theoretical ‘model’ is employed to explain improvements in quality, there are two reactions that proponents of greater use of information focus on: changes to provider behaviour and changes to patient behaviour.

### *Changing provider behaviour*

An increasing number of schemes further seek to encourage these changes by linking incentives with measurement of quality and publication of the results. In some cases this is quite explicit, as in pay-for-performance schemes that link some portion of physicians’ income to achievement of specific targets. The new GP contract is a prominent example of this.<sup>23,24</sup> But responses from healthcare providers to several publication schemes, where no financial incentive was offered, suggest an alternative model of “kudos and censure”. In this model, public disclosure of performance information places pressure on individual providers to improve their performance, regardless of explicit financial incentives or any

22. Marshall M, Shekelle P, Davies H, Smith P, Public Reporting on Quality in the United States and the United Kingdom, Health Affairs, 2003

23. Doran T, Fullwood C, Gravelle H, “Pay-for-Performance Programs in Family Practices in the UK”, New England Journal of Medicine, 2006

24. Lindemauer P, Remus D, Roman S, Rothberg M, Benjamin E, Ma A, Bratzler D, “Public Reporting and Pay for Performance in Hospital Quality Improvement”, New England Journal of Medicine, 2007

response from patients. After all, no-one wishes to be seen to perform badly.

“ A great hope of many who advocate publishing information about healthcare quality is that patients will choose the highest quality services, thus increasing the market share of high-quality providers and driving the poorest performers “up or out” ”

### *Changing patient behaviour*

A great hope of many who advocate publishing information about healthcare quality is that patients will choose the highest quality services, thus increasing the market share of high-quality providers and driving the poorest performers “up or out” (encouraging improvement in the face of lost business or forcing them to leave the market). Therefore, publication of performance information has often been expressed in terms of choice between different

providers.<sup>25</sup> However, as our review of patient responses to information demonstrates, the majority of patients do not identify “choice” as their preferred use of performance information, so this expectation may be optimistic.

An alternative model is one in which patients use performance information to strengthen their position in regard to health services, either by boosting their ability to challenge elements of their care that they do not understand or agree with, or by having a better understanding of how well their doctor or hospital provides care. Proponents argue that, for personal and practical reasons, patients will not readily change their provider. They suggest that the information should be used to improve the existing doctor-patient relationship rather than as a tool for seeking new providers.

The information we have  
The NHS collects a huge amount of data. But the Government and Department of Health currently focus on ‘wrong’ meas-

25. <http://www.nhs.uk/England/Choice/CompareHospitals.aspx?id=RQM01&id=RJ122&id=RJZ01&id=RJ701&cache=RQM01RJ122RJZ01RJ701st=cyh&sSearch=5LG> (Last accessed 03.05.07)

26. Populus interviewed a random sample of 1,004 adults in Britain aged 18+ by telephone between May 22nd & 23rd 2007. Interviews were conducted across the country and the results have been weighted to be representative of all British adults.

27. Lower social groups would find relative information about healthcare providers useful (71 per cent for groups DE; 65 per cent for AB) and would “boost their confidence” more (29 per cent for social groups DE; 16 per cent for AB)

28. Patients want an overall rating (57 per cent) more than detailed information (37 per cent) about providers and want information about practices and hospitals (67 per cent), not individual doctors (29 per cent).

In order to understand better how patients react to information – and to test the concept that the publication of data encourages patients to change their healthcare provider - we commissioned a survey earlier this year.<sup>26</sup> The results provide an insight into the ways that the public and patients react to the publication of data.

Patients want information to challenge or seek reassurance from their doctor (38 per cent). Some simply want to measure performance (17 per cent) but only a minority would use it to choose or change their provider (16 per cent and 13 per cent respectively). In other words, choice is not the primary use to which respondents would put information about quality. This finding challenges the assumption that the primary result of publishing such information will be to stimulate consumerism through choice and thereby improve quality. Rather, “better understanding” and “boosting confidence” are more commonly cited uses for information, suggesting that empowering the individual is more valued than helping them choose between providers. Strikingly, the most deprived socio-economic groups, and the oldest age group, most commonly cite these two uses.<sup>27</sup>

Unsurprisingly, we also found that there is a strong desire for information: 79 per cent of respondents - especially women and disadvantaged groups - want a ‘scorecard’ outlining their local surgery/hospital’s performance.

Respondents generally wanted information presented in aggregated forms (overall ratings for a service rather than individual measures, and for hospitals/practices rather than individual doctors) but they also wanted to compare hospitals with an expected level of performance, rather than have them ranked in a league table against each other.<sup>28</sup>

urements of performance, which fall into three categories:

- Failure: such as mortality, re-admission and re-operation. These measurements may be useful for some operations, such as heart surgery (although even here most operations do not end in such failure), but they are largely irrelevant to measuring the impact of the majority of the health service's day to day work (e.g. treating chronic conditions such as diabetes or asthma).
- Throughput (predominantly in acute care): such as waiting times. The measurement of waiting times is important, not least because shorter waits for treatment may improve outcomes. But measurements of throughput do not account for the quality of care provided.
- Proxy, aggregated quality measures: such as the Healthcare Commission's rating. The health service does produce 'high level' information on quality, but as we explain below, this 'level' of information is insufficient to drive necessary improvements to performance.<sup>29</sup>

There is some information on quality of care that is easily available on websites covering selected conditions and procedures, and the following analysis is based primarily on three major websites covering the UK:

- *The NHS in England* ([www.nhs.uk/England](http://www.nhs.uk/England)) is provided by NHS Connecting for Health, which was formed as an agency of the Department of Health on 1 April 2005 with the primary role of delivering the National Programme for IT. Its website gives information, by specialty and sub-specialty, on the availability of services and waiting times.
- *Dr Foster Intelligence* ([www.drfooster.co.uk](http://www.drfooster.co.uk)) is a public-private partnership (between the Information Centre for Health and Social Care and

Dr Foster Holdings LLP) that aims to improve the quality and efficiency of health and social care through better use of information. Its website was launched in February 2006 and gives information by trust and hospital, by specialty and sub-specialty, on indicators of structure, process and outcome.

- *The Healthcare Commission* ([www.healthcarecommission.org.uk](http://www.healthcarecommission.org.uk)) is the independent inspection body for the NHS and the independent healthcare sector. Its website provides a hospital "health check", a complex summary assessment of quality of services, the main elements of which are: core standards;<sup>30</sup> national targets;<sup>31</sup> and new national targets.<sup>32</sup> The Healthcare Commission also gives information from national surveys of staff and patients, at the level of the hospital; and national clinical audits by hospitals at sub-specialty level. (The Healthcare Commission was created as the successor to the Commission for Health Improvement (CHI) in 2004.)

A fourth website, [www.health.org.uk/qquip](http://www.health.org.uk/qquip), is a good source for international comparisons on quality.

There is a danger, after ten years of data collection and analysis, of (prematurely) concluding that the NHS has largely provided what is required for assessing quality of care in hospitals, and that it is now time to move on to the more challenging areas of primary and chronic care, and to the development of the electronic patient record. These are indeed important areas, and chronic care consumes a large (and increasing) share of healthcare resources. There is also evidence that management in the community – measured through a range of process and outcome measures – both prevents expensive hospital admission and keeps patients healthier. The quality and outcomes framework (QOF) has given us both some reasonable meas-

29. For example, the practice of giving overall ratings for individual hospital trusts, which has been followed by the Government since 2000, has led to simplistic methods of assessment and a lack of information about lower organisational levels.

30. For safety, clinical and cost effectiveness, governance, patient focus, accessible and responsive care, care environment and amenities, Public Health.

31. These are dominated by waiting times for various services, but also include proportions of people suffering from a heart attack who receive thrombolysis within 60 minutes of calling for professional help; booking and choice of hospital admissions; and reorganising cancelled operations.

32. This heterogeneous mix of Government priorities includes: mortality rates from heart disease, stroke, suicides; methicillin resistant *Staphylococcus aureus* (MRSA) levels; participation of problem drug users in drug treatment programmes, health inequalities, adult smoking rates; under-18 conception rate; NHS patient experience; obesity among children under 11; 18-week wait from GP referral to hospital treatment; health outcomes for people with long term conditions; and reducing emergency bed days



ures of success and a consistent and high quality data flow. Despite much progress and effort we are still far from resolving the weaknesses in measuring the quality of hospital care.

Table 2, below, summarises the information we have on hospital admissions. This shows massive gaps in the information that is available once we go below the level of the hospital. Dr Foster Intelligence has helped to fill this gap by giving details for different hospital services. The best source of information on the clinical quality of care is provided by national clinical audits like those of cardiac surgery and stroke. However, information from national clinical audits is not easily accessible, and Dr Foster Intelligence does not report the details covered by these audits. These are, moreover, incomplete in that they omit the patient's experience.

The national patient survey programme of the NHS gives information on patients' experience of the processes of care, and indeed is the largest survey of its kind globally, but still cannot give information at specialty level. Perhaps most alarmingly, there is virtually no analysis of patients' experience of the outcomes of care, yet a tried and tested tool exists to allow this to be done, and BUPA has shown how we can collect this data routinely and act upon it.

The information we need

We consider information on four dimensions and from two perspectives. The four dimensions are: waiting times, structure, process, and outcome. Clearly, other dimensions could have been chosen, but these represent the overriding NHS priorities of the last ten years, and the most com-

**Table 2: available information on quality of hospital care from the NHS**

	Hospital	Specialty	Condition
Waiting times	Yes	Subspecialty	None
Structure	Some	Some	Selected
<i>Patient experience</i>			
Process	Some	Some	None
Outcomes	None	None	None
<i>Clinical</i>			
Process	Health check	Health check Some national clinical audits	Health check Some national clinical audits
Outcomes	Mortality and readmission rates	Mortality and readmission rates, some national clinical audits	Mortality and readmission rates, some national clinical audits

mon areas for performance assessment and improvement. The two perspectives are the clinician’s and the patient’s. There is no need to explain what we mean by waiting times, but the other terms do require clarification. (The elements of structure, process, and outcome were identified as being fundamental to understanding quality of care in the seminal work of Donabedian.)<sup>33</sup>

- *Measurements of structure.* These include, for example, the adequacy of facilities and equipment, the qualifications of medical staff, and administrative structure. These measurements need to be used sparingly. They have limitations as they can encourage a “tick-box” mentality. That is, providers can become preoccupied with getting the right committees in place without changing the way anything is provided. However, evidence-based structural measures can be invaluable. For example, there is good evidence that patients treated in specialist stroke units fare better than in non-specialist hospital units.

- *Measurements of clinical process.* These use evidence of relationships between outcomes and the processes of care. The Government’s national service frameworks (NSF) and guidelines recommend what processes of care ought to be used for each condition.
- *Measurements of clinical outcomes.* These include mortality and hospital readmission rates. Indeed, these are the only data that the NHS routinely collects following discharge from hospital.
- *Measurements of process from the patient’s perspective.* These include quality of food, accommodation, cleanliness and being treated with dignity and respect.
- *Measurements of outcome from the patient’s perspective.* These capture data other than whether the patient died or was readmitted: in particular, whether that patient’s state of health has improved following discharge from hospital.

Table 3 summarises these perspectives and dimensions.

**Table 3: information the NHS needs**

	Clinical perspective	Patient perspective
Waiting times	Self explanatory	
Structure	The adequacy of facilities and equipment, the qualifications of medical staff, administrative structure	
Process	Measures of clinical process – application of evidence-based care, e.g. following NSF and NICE guidelines	Quality of food, accommodation, cleanliness and being treated with dignity and respect
Outcome	Include mortality and hospital readmission rates	Whether that patient’s state of health has improved following discharge from hospital

33. Donabedian A, "Evaluating the Quality of Medical Care", The Milbank Quarterly, 2005

### Obstacles to the use of healthcare outcome measurements

Despite enthusiasm and effort the NHS, like most health services globally, has made little progress in measuring its performance. The barriers to successful measurement (complexity, credibility, level, technology, technical expertise, organisational cost and 'primary care') are described in detail below:

- *Complexity:* Measuring outcomes in a sophisticated and accurate manner is difficult, while using simple output measurements like case fatality can be misleading. Many different measurements are currently available and no consensus has been reached on which should be preferred, or about which should take priority when they conflict.<sup>34</sup> Conceptual problems, explained in further detail in chapter five, include being able to adjust for the case-mix (the different level of risk offered by each individual in a group of patients) of patients being treated. This is similar to the arguments about the need to estimate the value added by schools to take account of the different mix of children on entry. Highly skilled surgeons who operate on difficult cases may have worse outcomes than less skilled surgeons who operate on simple cases. There are also problems in deciding the period during which outcomes ought to be measured.
- *Credibility:* Outcome measurements that exist are often not seen as credible by health professionals. This may on occasion be due to a fear of having their professional independence challenged or being held accountable for their performance. In contrast, where measures are sufficiently sensitive and credible, for example the approach taken to the publication of cardiac surgery mortality figures, measurement gets relatively high levels of support from clinicians.
- *Level:* A further problem is getting the 'organisational level' of the information right. There are two ways of simplifying the assessment of the quality of a hospital, and each aims to give a global indicator of hospital performance and hence of all its services. First, one can assume that the part of a hospital's services that we can measure will act as a good proxy for all services in the hospital. For example, the outcome measurements we have for mortality and readmission rates may be assumed to give a good indication of the quality of services generally. Second, one can make estimates of the quality of the services provided at hospital level by measuring the satisfaction of patients with the processes of care, such as catering, cleaning, and the extent to which they were treated with dignity and respect.

Unfortunately, these methods are too blunt. A general acute hospital is a highly complex organisation. The CHI, in its inspections of the implementation of systems and processes to assure and improve the quality of care in acute hospitals, found that single-specialty hospitals tended to do best. Performance in multi-specialty hospitals varied greatly and there was often a dysfunctional clinical team.

In the case of multi-specialty hospitals, generic indicators are valid for those services that are organised at hospital level (catering, cleaning, diagnostic services), but not for those that vary between specialties and individual doctors. Thus, to know that a hospital has low mortality rates for medical and surgical admissions tells us nothing about mortality rates in surgical sub-specialties. In the Bristol case, the numbers of excess deaths from paediatric cardiac surgery would have been too small to have had an impact on its total rates of mortality; and there were important

34. Such as service costs, end to end response times, capacity levels, demand, referral rates, numbers of missed appointments ('DNAs'), clinical outcomes, waiting times/numbers by health professional/referring doctor specialty, length of stay, patient satisfaction.

differences in outcomes between the two surgeons for different procedures.

- *Technology*: Poor technological infrastructure has hindered the use of appropriate outcome measures. This is an issue that the Connecting for Health programme has attempted to address.<sup>35</sup> For cancer care, there are registries that provide a history of treatment for individuals. There is hope that electronic records will do this in the future for all patients routinely, but the record of successful implementation of such massive IT projects cautions against this being a practical reality in the next few years.
- *Technical expertise*: despite the creation of the Health and Social Care Information Centre in April 2005<sup>36</sup> there remains a shortage of staff, particularly at a local level, who are able to collate and analyse outcomes data. CHI consistently found that use of information within NHS trusts was the least developed of the various components of clinical governance.

- *Cost*: a key challenge for outcomes measurement is to ensure that the cost of collecting data and ensuring completeness, accuracy and standardisation are justified by the benefits derived. This assurance has not been proven.
- *Primary Care*: It is much more complex to assess the quality of primary care. Chronic care, which accounts for much of primary care, usually involves multiple providers, multiple interventions and, by definition, sees the patient's health outcome decline over time.

As we show in the case studies that follow, evidence indicates that outcome data can drive improvement in the quality of services, such as in cardiac care where it has helped improve the quality of work performed by surgeons.

The next chapter gives examples of when information has been instrumental in improving services, and when it has been ineffective.

35. Formally the National Programme for Information Technology.

36. <http://www.ic.nhs.uk/services/> (Last accessed 03.05.07)

---

# 3

## From information to action

### Five case studies

Plausible models, supported by good research, suggest that measuring quality of care and publishing the resulting information can improve healthcare. But it is clear that large gaps in the available data remain, and that there are considerable technical challenges to moving forward. To identify some potential ways of solving these problems we now consider five case studies of attempts to use information to improve quality (three from the US,<sup>37</sup> two from England and one from Scotland):

- The publication of mortality rates following coronary artery bypass grafting (CABG) surgery in the Cardiac Surgery Reporting System, which began in New York state in 1989;
- The publication of annual reports of clinical outcome indicators by Scotland's Clinical Resource and Audit Group (CRAG), 1994-2002, for use by the public and doctors;
- The Veterans Health Administration National Surgical Quality Improvement Program (NSQIP), which began in 1994;
- The annual star rating of the NHS in England, (2001-2005), a single aggregate score (zero to three) linked to sanctions and rewards;
- The London Patient Choice Project (LPCP), which developed systems to give patients who had been waiting for more than six months the option of

choosing another hospital to have their operation sooner.

Each of these uses of information is examined for purpose, selection of indicators, methods, presentation, dissemination and impact. Analyses typically find large variations in quality, whatever the measure (processes or outcomes). It is wrong to assume that, where clinical care has not yet been measured, there is no variation in quality. Ignorance ought not to be bliss. The significant gaps in the information that is available about NHS hospitals are a serious cause for alarm.

#### Cardiac surgery reporting system in New York State

##### *Summary*

- In New York, outcome measurement and publication encouraged both quality improvement and appropriate changes to the supply side of the market. Publication was followed by improvements in outcome, and stimulated some quality improvement activity. Whether from “knightly” or “knaveish” considerations, poorly performing surgeons improved performance or ceased operating altogether.
- There is also some evidence of gaming, such as outmigration and upcoding of risk, at least at the margins. The noted changes in behaviour were by providers rather than patients.

37. Appendix one provides an overview of the American Healthcare system, and why, despite its flaws, we believe that we can learn from the US in this particular case.

- As there is no effective ‘control’ in the New York studies, we do not know the relative importance of measurement and publication. Northern New England states achieved, through private reporting, similar results to New York, so it may be that measurement alone is sometimes enough.

### *Purpose*

Perhaps the most famous and studied attempt to use information to improve quality of care began in 1989, when the New York State Department of Health developed methods to collect and analyse data to reduce mortality after CABG.

### *Selection*

The focus on CABG surgery represented a conscious break from previous attempts at performance measurement, such as the Healthcare Financing Administration annual reports on mortality among hospitalised Medicare<sup>38</sup> patients. The differences were twofold: the scheme was focused, and it was, or attempted to be, risk-adjusted. Cardiac surgery is a definite, singular intervention (an operation), for which there is an outcome (death or survival), which has a clear, if not necessarily causal, relationship with the intervention. This contrasts with mortality rates for a whole hospital, where a range of interventions, confounding factors and patient case-mix influence the results.

Much of this work was led by the cardiac advisory committee of the state’s Department of Health: a group of cardiac surgeons, general physicians and consumers who guide the department on cardiac care. To adjust for the relative risk that different patients presented, the department developed the Cardiac Surgery Reporting System (CSRS), whereby clinical data was collected on all patients undergoing cardiac surgery in New York, such as their risk factors and demographic details. Risk factors included unstable angina, low ejection fraction, stenosis of the left main

coronary artery, congestive heart failure and chronic obstructive pulmonary disease.

Hospitals provided the data on a quarterly basis to the department, which then, under the guidance of the committee, compared the mortality rates of the providers. This comparative data included crude, expected and risk-adjusted 30-day mortality rates by hospital and surgeon was given to the providers and became the basis for efforts at improving quality of care. As even a critical study of the New York experiment observed: “With these important improvements, CSRS became the first profiling system with sufficient clinical detail to generate credible comparisons of providers’ outcomes. For this reason, CSRS has been recognized by many states and purchasers of care as the gold standard among systems of its kind.”<sup>39</sup>

### *Dissemination*

Initially, the department released only the data on hospitals to the general public, reserving the data on specific surgeons solely for the providers. However, a lawsuit by the newspaper *Newsday* under the state’s Freedom of Information laws led to the general publication of data on surgeons in December 1991. This provoked tremendous hostility from physicians, including the committee, which felt that the number of operations performed by each surgeon was too low to be statistically significant. The committee responded by recommending that hospitals submit data that would make it impossible to detect the performance of individual surgeons. A compromise was eventually reached. Data would only be released about surgeons who had performed at least 200 operations during the preceding three-year period.

### *Model of improvement*

As originally envisaged (that is, with confidential reporting) the scheme would have made an appeal to the professionalism and

38. A federal health insurance programme for people age 65 and over and for individuals with disabilities.

39. Green J, and Winfield N, “Report Cards on Cardiac Surgeons - assessing New York state’s approach”, 1995

altruism of individual surgeons. Following the intervention of *Newsday* and the publication of the data, the model of improvement became an implicit incentive for providers, for there was now a threat to their reputation and market share.

### Results

Mark Chassin, commissioner of the department from 1992 to 1994, argues that “the registry produces reliable and valid measures of quality, and hospitals and cardiac-surgery programs throughout New York use this information to improve outcomes for their patients.”<sup>40</sup> The evidence is that a remarkable improvement in outcomes for cardiac surgery followed the introduction of the reporting system. From the beginning of 1989 to the end of 1992, risk-adjusted mortality for CABG in New York fell by 41 per cent, going from 4.17 per cent to 2.45 per cent.<sup>41</sup> This was a far more dramatic decline than was achieved elsewhere: Medicare patients undergoing CABG, for example, enjoyed only an 18 per cent reduction in mortality nationwide between 1987 and 1990.<sup>42</sup> Only in northern New England states were similar results achieved during this period – and perhaps significantly, these states were subject to their own outcomes reporting process, if only a confidential one.

### Impact

Despite the impressive results achieved in New York, publication of CABG mortality rates had no impact on the flows of patients to hospitals: improvements reflected changes in the behaviour of providers rather than patients. The reported improvements followed from three kinds of actions by providers:

- Stopping surgeons with poor outcomes from undertaking CABG. Individual surgeons with bad mortality rates left the market.<sup>43</sup> More generally, the results showed that surgeons who performed

fewer than 50 operations per year had higher mortality rates than high-volume surgeons. Some hospitals, therefore, began to restrict the operating privileges of low-volume surgeons: 27 stopped performing bypass surgery altogether between 1989 and 1992. All of these had had mortality rates in their final year of practice that were between 2.5 and 5 times higher than the state average.<sup>44</sup> Other surgeons whose performance was unsatisfactory (many of whom had received their principal training in fields other than adult cardiac surgery, such as paediatric surgery) chose to retire from cardiac surgery. As poorly performing surgeons were moved away from the operating theatre, high-performers were given more high-risk patients.

- Poorly performing providers implemented specific quality improvement reforms. The committee intervened to provide expert advice to poor performers. In one case, a hospital suspended cardiac surgery until a new chief could be installed to reform the entire department.<sup>45</sup>
- Poorly performing providers also sought to understand the precise causes of their high mortality rates. Where these causes were not obvious in the initial data, they undertook research. Dziuban et al cite the example of St Peter’s Hospital in Albany, the state capital, which struggled to understand why its mortality rate was significantly higher than the state average.<sup>46</sup> Initial studies of mortality and morbidity showed no problems with quality of care, and reviews of risk-factor coding exposed no faults in the data. Eventually, with the help of the department, the hospital pinpointed its emergency provision of bypass surgery as the source of the problem: the mortality rate there was 26 per cent, compared to a state average of 7 per cent. A further review concluded that the problem was

40. Chassin M et al, “Benefits and Hazards of Reporting Medical Outcomes Publicly”, *New England Journal of Medicine*, 1996

41. Hannan E et al, “Improving the outcomes of coronary artery bypass surgery in New York State”, 1994

42. Peterson E et al, “Changes in mortality after myocardial revascularization in the elderly: the national Medicare experience”, *Annals of Internal Medicine*, 1994

43. Chassin et al 1996

44. Ibid

45. Ibid

46. Dziuban S et al, “How a New York cardiac surgery program uses outcomes data”, *Annals of Thoracic Surgery*, 1996

the failure to stabilise patients adequately before surgery (not using intra-aortic balloon pumping as much as required, for example). Once this specific failing had been addressed, St Peter's mortality rate improved to normal levels.

Omoigui et al raised questions over the various types of gaming behaviour undermining the reported successes. They suggested that the 41 per cent reduction in risk-adjusted mortality from 1989 to 1992 did not mean a genuine improvement in outcomes. First, a number of high-risk patients were treated in a neighbouring state (Ohio).<sup>47</sup> Other studies have, however, questioned the scale of this outmigration.<sup>48</sup> Moreover, as Omoigui et al recognised, it is not possible to ensure that all providers accept high-risk patients in a system where outcome measures are used nationwide. Their second criticism was that (despite outmigration), there was a dramatic increase in the reporting of risk factors of patients undergoing CABG after 1989: the expected mortality using the reported risk rose from 2.62 per cent in 1989 to 3.54 per cent in 1992. The authors stated that "since surgeons and hospitals are not only aware that they are under surveillance but that they are also responsible for primary data collection, artificial increases in patient severity scores could result from the selective emphasis of clinical characteristics".

If a system is undermined by gaming this does not necessarily mean that it is fatally flawed. Given the discussion on "knights and knaves", any system that aims to drive improvement will be likely to produce the "knaveish" response of gaming. Indeed, we suggest that an absence of gaming indicates an absence of incentives for improvement. The above account shows that the New York reporting system produced data that surgeons and providers found credible. They were very much aware of the information, as this was in the

public domain, even though it had no discernible impact on patients' behaviour. The information was published so that it was easy to interpret. It was used in benchmarking, and was reported at the organisational level where corrective action could be taken.

## Clinical Resource and Audit Group Outcome Indicators Reports

### Summary

- The Clinical Resource and Audit Group (CRAG) carried out a pioneering study that attempted, largely unsuccessfully, to change provider behaviour.
- The two evaluations of the CRAG Outcome Indicators reports provided criteria on which those reports failed: credibility, awareness, ease of interpretation, reporting results at the levels where corrective action can be taken, providing information in a form that can be used for benchmarking, timeliness, and use of incentives for providers.
- As Florence Nightingale constantly reminded herself, recommendations have to be carried out: there must be incentives for providers to act on what has been reported.

### Purpose

In 1994, Scotland's Clinical Resource and Audit Group (CRAG) pioneered in Europe the publication of annual reports of clinical outcome indicators for use by the public and doctors.<sup>49</sup> The group's reports aimed to provide a benchmarking service for clinical staff by publishing comparative clinical outcome indicators across Scotland. Performance management was seen as a matter for local action: there were no external national pressures to act on CRAG's indicators other than publication.

### Selection

CRAG reports included two kinds of hospital clinical indicators: emergency read-

47. After reviewing 9,442 CABG operations performed in neighbouring Ohio between 1989 and 1993, the authors found that patients referred from New York were more likely to have had prior open heart surgery than patients from Ohio or other states or other countries (44 per cent compared to 21.5 per cent and 37.4 per cent and 17.3 per cent respectively). They were also more likely to be in the New York Health Heart Association's functional class III or IV, another sign of high risk. Both the expected and the observed mortality rates for the patients from New York were higher than for the other referral cohorts, a pattern that was not evident in the control period 1980-1988.

48. Chassin M, Hannan E, DeBuono B, "Benefits and Hazards of Reporting Medical Outcomes Publicly", *New England Journal of Medicine*, 1996

49. "Clinical Outcome Indicators. A report from the Clinical Outcomes Working Group", CRAG, 2002



mission rates for medical and surgical patients; and mortality (or survival) after hospital treatment for hip fracture, acute myocardial infarction, stroke and selected elective surgery.

### *Methods*

An important development was the linking of data on hospital admissions and discharges to mortality data collected by the Registrar General, so that indicators of mortality (or survival) took account of deaths both inside and outside hospital. These indicators were standardised for age, sex and deprivation (based on the patient's area of residence) but not for risk (unlike New York's CSRS). The reports emphasised that variations in reported performance could have been due to varying proportions of higher and lower risk procedures, differences in the condition of patients on admission and differing diagnostic thresholds. CRAG reported performance on each indicator using methods deemed appropriate for each.

### *Presentation*

The last report presented information on the following:

- survival (at 30 days and 120 days) after hospital admission following hip fractures, acute myocardial infarction and stroke: total numbers of patients and those who survived, as well as crude and standardised mortality rates, and ten-year trends in survival for single years (without confidence interval estimates).
- surgical mortality (at 30 days) over three three-year periods: numbers of operations and deaths, and crude and standardised mortality rates (with 95 per cent confidence interval estimates).
- readmission rates (at seven and 28 days): total numbers of discharges and readmissions, and crude and standardised readmission rates.

### *Dissemination*

Initially, publication attracted the attention of the media (results were sometimes presented as "death league tables"). Over time, however, media interest waned and information was disseminated exclusively to chief executives and senior clinicians. The overriding concern of the last report in 2002 was to highlight the limitations of these indicators. The report included the results of an investigation into the wide discrepancy found in survival rates after admission for stroke between two hospitals in Edinburgh. The explanations offered were that the closure of the Accident and Emergency unit at one of the hospitals meant that severe emergency cases of stroke were more likely to be taken to the other hospital, where the patient's first milder stroke may not have been recorded.<sup>50</sup> An outsider is struck by the note of satisfaction which is evident in the way this negative conclusion was reported.<sup>51</sup>

### *Model of Improvement*

Given the lack of widespread dissemination, and the markedly lower enthusiasm for encouraging choice in provision of healthcare in Scotland than in England, the CRAG reports have primarily operated by changing provider behaviour; they have not used any incentives beyond appeals to the intrinsic motivations of professionalism and altruism.

### *Impact*

There have been two evaluations of CRAG reports. One was by independent academics based in England,<sup>52</sup> who explored their effects on NHS trusts in Scotland. The other was by a CRAG-funded clinical indicators support team,<sup>53</sup> which investigated the requirements of health boards and trusts for clinical performance information. They checked whether the indicators met those requirements, and how and why they had been used. The main conclusion of the academics, Mannion and Goddard,

50. The report's concluding observation on this investigation is: "This is just one example (albeit perhaps an extreme one) of the kind of issues relating to case mix which may affect apparent outcome in the context of this indicator of survival after admission with stroke".

51. *Ibid*

52. Mannion R, and Goddard M, "Impact of published clinical outcomes data: Case study in NHS hospital trusts", *British Medical Journal*, 2001

53. "Clinical Outcome Indicators. A report from the Clinical Outcomes Working Group", CRAG, 2002

was that in Scottish trusts these indicators “had a low profile...and were rarely cited as informing internal quality improvement or used externally to identify best practice”.<sup>54</sup> The clinical indicators support team came to similarly depressing conclusions.

Unlike New York’s cardiac surgery reporting, clinical staff did not consider the information published to be credible because of the limited accuracy of coding and the lack of an adjustment for risk. Also the information was seen as being out of date (at least 18 months old at the time of publication). There was low awareness of the reports because of lack of media attention and limited dissemination within trusts. The indicators were neither linked to a formal system of performance assessment nor used by health boards in holding trusts to account.

The two evaluations emphasise what the reports did *not* do. They did not identify: good or poor performance; key messages; where action could be required; how services were organised where outcomes were good; or possible reasons for variations in outcomes. Clinicians did not find the information useful for benchmarking because it was presented at the level of the hospital rather than the clinical team.

In many ways, the CRAG study illustrates how measurement alone will not improve quality. The next case study we investigate shows how measurement allied to a comprehensive quality improvement programmes can have significant impact.

#### Veterans Health Administration National Surgical Quality Improvement Program

##### Summary

- The Veterans Health Administration (VA) is widely recognised as having transformed itself and become a health service provider of very high quality in the last 15 years. This has been

achieved, in part, through a quality improvement programme founded on the collection and internal reporting of process and outcome measurements.

- In many senses, the VA is more like the NHS than many other parts of the US healthcare system: it provides care for a defined population, it is a system that integrates primary and secondary care, and it is a Government provided service. As such, its approach eschewed market mechanisms (its population of patients are “captive”) and may be more easily transferable to the UK than others in the US.
- The VA has powerful lessons to teach about how to use information to improve services. The following is required to ensure that information is acted upon to improve services:
  - a formalised, non-punitive, feedback on performance;
  - sharing of best practice between different organisations;
  - using outlying performance as a spur to investigate, identify and correct underlying problems.
- This approach is one that relies on *internal* rather than *public* reporting. It is the approach we label “information for improvement” in our vision for 2010. (See chapter 7).

##### Purpose

The Veterans Health Administration, a part of the Department of Veterans Affairs (commonly known as the VA), is the largest provider of healthcare in the US, and is aimed at citizens who have served in its armed forces. As of 2002, it had 159 medical centres, of which 128 performed major surgery, 165 long-term care facilities and 376 outpatient clinics.<sup>55</sup> The VA is the closest thing in US healthcare to the National Health Service. For many years, the VA was held up as an awful example of the perils of socialised medicine, as its quality of care was notoriously poor. Studies in the 1980s

54. Mannion R, and Goddard M, 2001

55. Khuri S, Daley J and Henderson W, “The comparative assessment and improvement of quality of surgical care in the department of veterans affairs”, *Archives of Surgery*, 2002

and early 1990s showed that the VA provided lower standards of care than private providers funded through the federal Medicare programme.<sup>56</sup>

In 1986, Congress passed legislation requiring the VA to report annually the outcomes of its surgical operations compared to the national average, and to adjust those outcomes to take account of patients' conditions. The VA surgeons who were commissioned to respond to the congressional mandate reported that there were neither any known national averages for surgical outcomes nor any known risk-adjustment techniques that could be used for the various types of surgery.

There are a number of advantages for the VA in being a NHS-type system. It has:

- centralised authority;
- capacity for sophisticated medical informatics to develop both national averages and risk adjustment techniques, and hence measure in a systematic and reliable way the quality of care it provides;
- a defined and captive population;
- the ability to drive change through internal processes, such as private reporting and intensive quality improvement efforts (rather than open publication of outcomes that threaten the loss of market share through competition).

The purpose of the VA programme was to collect the information necessary for a comprehensive system of quality improvement.

### Selection

The VA began by collecting data on outcomes of surgery, but then extended this to different types of care:

- *Preventive*: mammography, influenza vaccination, cervical cancer screening;
- *Inpatient*: providing aspirin and beta blockers at discharge for acute myocardial infarction;

- *Outpatient*: screening for diabetes by glycosylated haemoglobin measurement and annual eye examination, and screening for depression.

### Methods

The VA's first step was to launch a study of surgical risk. This used data from 117,000 major operations collected by a nurse in each of the 44 VA medical centres between 1991 and 1993. The study developed and validated risk-adjustment systems for the measurement of mortality and morbidity in nine different types of surgery (including eight non-cardiac ones) 30 days after the operation. It found that health outcomes were determined jointly by the risk factor of patients before surgery, random variation and quality of care. If the first two could be accounted for by appropriate statistical methods, it would be possible to regard health outcome as a measure of quality of care.

Satisfied with the validity of these techniques, the VA was ready to proceed with a programme to monitor the quality of care in all VA medical centres performing major surgery. This program (NSQIP) began in 1994. A designated nurse at each centre collected data about 30-day mortality and morbidity and sent it for analysis to one of two centres, one for cardiac surgery (the Centre for Continuous Improvement in Cardiac Surgery), the other for non-cardiac surgery (the Hines VA Co-operative Studies Program Coordinating Centre). By the end of fiscal year 2000, there were 727,447 operations recorded in the NSQIP database.<sup>57</sup> These centres edit the data and refer any potential errors back to the reporting centre for resolution.

These outcomes data were supplemented by process data that measured what proportion of interventions were the correct ones for patients' circumstances. For example, in the area of preventive care, were influenza vaccinations

56. Gardner J "VA on the spot: care quality to be probed by Congress", *Modern Healthcare*, 1998. Also Kent C, "Perspectives: VA under fire for quality of care", *Faulkner & Gray Medical Health*, 1991, and Zook C. et al, "Repeated hospitalisation for the same disease: a multiplier of national health costs", *Millbank Memorial Fund Quarterly Health Soc*, 1998

57. Khuri et al, 2002

provided? In inpatient care, was aspirin provided to a patient at discharge after acute myocardial infarction? Using outcomes and process data together would serve as a meaningful measure of the quality of care, and allow quality improvement to be targeted at the precise elements of care that are failing.

NSQIP does not gather data about specific surgeons. There are two reasons for this emphasis on systems rather than individuals.<sup>58</sup> The manifest justification is that “if systems of care are poor at a specific institution, even the most competent surgeon can have poor outcomes. Likewise, a mediocre surgeon can have excellent outcomes if he or she is functioning in an environment with excellent systems of care”.<sup>59</sup> A latent justification was the need to win the support and trust of surgeons in the whole reform process. There were concerns that too much scrutiny of individual surgeons would alienate them, when their co-operation was vital to the collection of data and the processes of quality improvement.

### *Dissemination*

The guiding principle of the programme is to effect quality improvement by providing feedback to VA centres (and not by publication). As Khuri, Daley and Henderson observe, “the NSQIP is not a punitive program for the purpose of identifying ‘bad apples’. On the contrary, its primary focus is to provide the surgeons and managers in the field with reliable information, benchmarks, and consultative advice that will guide them in assessing and continually improving their local processes and structures of care”.<sup>60</sup> The feedback takes the form of:<sup>61</sup>

- *an annual report to each centre.* The report includes risk-adjusted outcomes data for all participating centres, though they are given anonymous codes rather than identified by name. Each centre knows only its own code,

which allows comparison of their performance with the other centres.

- *assessment of “high and low outlier” centres.* At an annual two-day meeting the executive committee reviews the performance of each centre over the previous four years. The committee reports concerns over “high outliers” (centres whose rates of deaths or complications are much higher than expected) and communicates praise and certificates of commendation to “low outliers” (centres whose observed rates of deaths and complications are much lower than expected). This can be done not only for whole centres but, thanks to the detail of the information, also to surgical specialties within each centre.
- *spreading best practice:* centres that have been consistently low outliers, or which have made dramatic improvements, are encouraged to reveal their methods. This information is included in its annual reports to the centres.
- *site visits:* providers can invite the programme to conduct structured site visits to identify and propose solutions to problems with care provision. The visit has two parts. First, a NSQIP nurse checks whether the data collected at the site is reliable. If it is, then a team consisting of a surgeon, a critical care nurse, an anaesthetist, and a health services specialist arrives and provides senior surgeons and managers at the centre with a confidential report.
- *provision of self-assessment tools:* providers and managers are given instruments developed by NSQIP to assess the strengths and weaknesses of their work, especially when the reports show them to be high outliers.

### *Model of improvement*

The VA is perhaps the finest example of information being used as part of an internal improvement agenda, yet there

58. Young G, Charns M, Daley J et al, “Best practice for managing surgical services: the role of co-ordination”, *Journal of Healthcare Management Review*, 22, 1997. Also Young G, Charns M, “Patterns of co-ordination and clinical outcomes: a study of surgical services”, *Health Services Research*, 1998

59. Khuri et al, 2002

60. *Ibid*

61. *Ibid*

was no risk of loss of reputation (at least in public) or market share, and no hint of consumerism (the VA has a captive market).

**Impact**

The improvement in the quality of care achieved by the VA after implementing the systems detailed above:<sup>62</sup>

- Since the National Surgical Quality Improvement Program began collecting data in 1991, 30-day mortality in major surgery at VA centres has fallen by 27 per cent, while 30-day morbidity has fallen by 45 per cent.<sup>63</sup> And, unlike in New York State, there was no change in the risk profiles of patients during this period.
- A study by Jha et al showed that, before the implementation of the programme in the baseline period (1994-95), the VA's performance was poor in nearly all the areas assessed. Measurements taken

after its implementation (1997 to 2000) showed significant improvement in preventive, inpatient and outpatient care (Table 4). Of the 13 types of care for which multi-year data was available, there were significant improvements in 12 of them.<sup>64</sup>

- A comparison of the VA's results with those of Medicare for the years 1997 to 2001 showed that the VA outperformed Medicare in every type of care in 1997-99, and in 12 of 13 types of care in 2000-01 (the exception was annual eye examinations for diabetes patients). Jha et al concluded that the only plausible explanation is the implementation of NSQIP, as nothing else had changed. They also argued that these comparisons underestimated the VA's achievements, since the VA's patients are generally sicker and more likely to be physically disabled, mentally unwell, poor, uneducated and from a disadvantaged ethnic group than the general population.<sup>65</sup>

**Table 4: improvements by the VA in preventive, inpatient and outpatient care (%)**

Type of care	1994-95	1997	1998	1999	2000
<i>Preventive</i>					
Mammography	64	87	89	91	90
Influenza vaccination	28	61	67	75	78
Cervical cancer screening	62	90	93	94	93
<i>Inpatient</i>					
Providing aspirin at discharge for acute myocardial infarction	89	92	95	97	98
Providing beta blockers at discharge for acute myocardial infarction	70	83	93	94	95
<i>Outpatient screening</i>					
Glycosylated haemoglobin measurement (diabetes)	51	84	90	94	94
Annual eye examination (diabetes)	48	69	72	73	67
Depression	-	-	44	62	73

62. In April 2007 the McClatchy Newspaper Group challenged the VA success story. The heart of the story is that some VA press releases may have oversold the studies' impact and there is some anecdotal evidence of marginal gaming, particularly on waiting times. In our view there is nothing of any substance to challenge the VA results.

63. Khuri et al, 2002

64. Jha et al, "Effect of the Transformation of the Veterans Affairs Healthcare System on the Quality of Care", New England Journal of Medicine, 2003

65. Ibid

Another advantage of NSQIP is the evidence that it provides on outcomes. The New York programme found a strong link between volume and outcomes for bypass surgery and the VA had a policy of closing surgeries that handled a low volume of cases. But one of its studies found that what applied to CABG surgery could not be generalised to other types of surgery. The VA consequently reversed the policy, which would otherwise have reduced access.<sup>66</sup>

The VA programme of surgical improvement started at about the same time as Scotland's CRAG reports. While the former has had a dramatic impact, the latter seems to have had none. Jha et al in their explanation of the VA's success make it clear how it differed from CRAG in awareness, credibility, relevance and importance:

- performance contracts held managers accountable for meeting improvement goals;
- whenever possible, quality indicators were designed to be similar to performance measures commonly used in the private sector;
- data gathering and monitoring were performed by an independent agency – the external peer review program;
- critical process improvements, such as an integrated, comprehensive electronic medical record system, were instituted at all VA medical centres;
- performance data was widely distributed within the VA, among key stakeholders such as veterans' service organisations, and among members of Congress.<sup>67</sup>

#### A transferable model?

The success of the VA's programme is clear, but can it be applied in England and Wales? The parallels between the organisational arrangements of the VA and the NHS – such as their highly cen-

tralised structures of decision making and their ability to run a unified information system – are greater than between the NHS and any other element of US healthcare. This may mean that the UK is uniquely well-placed to copy and benefit from the VA's reforms. The model used by the VA relies heavily on internal kudos and censure, with the information held within the system rather than being more publicly reported. While VA beneficiaries may be able to get access to such information, the more general public cannot. This may be reasonable given the limited population it serves, but the NHS cannot be so parsimonious with its data.

Given the plurality of healthcare provision in the US, transferability can be to some extent judged by how successfully the VA model has been adopted by US private sector providers.<sup>68</sup> The signs here are encouraging.

In 1999, a private sector initiative involving three non-VA medical centres (Emory University Hospital, the University of Kentucky Chandler Medical Center, and the University of Michigan Medical Center) was launched to discover whether the VA's data-gathering methods could be used by non-VA providers, and whether the models it used to predict outcomes from surgery could be applied to non-VA patients.<sup>69</sup> A nurse from each of these non-VA centres collected and provided data to NSQIP Data Coordinating Centre at Hines. Analysis of the first year's worth of data showed that NSQIP data collection methods were applicable to all three non-VA centres, and that the VA risk adjustment methods were applicable to non-VA patients. Indeed, the results were so encouraging that the NSQIP executive committee aimed to make NSQIP available to the surgical community across the US.

If the VA model can be adopted by other providers in the US, there is no

66. Kuhri S, Daley J, Henderson W et al, "The relationship of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program", *Annals of Surgery*, 1999

67. Jha et al, "Effect of the Transformation of the Veterans Affairs Healthcare System on the Quality of Care", *New England Journal of Medicine*, 2003

68. Khuri et al, 2002

69. Khuri et al, 2002

reason to suppose that the NHS – which in many ways is closer to the VA health system – cannot do likewise.

### Criticisms

Yet, there are still reasons for caution. In the VA programme the more useful information at doctor or hospital level is not made available in a simple report card for public use. This lack of openness is quickly becoming politically unacceptable in the NHS.

In addition, the range of measurements is relatively limited. The findings of the Jha study were striking, but the indicators it looked at were process measures rather than outcome measures. Whether a patient has been vaccinated, for example, is a different question from whether and to what extent a patient's health has actually improved as the result of a medical intervention. There is a strong link between process and outcome (a patient who has received all the appropriate types of care during a course of treatment is likely to enjoy better health than one who hasn't) but, as the authors note, "A full assessment would require the measurement of outcomes such as...patient satisfaction."<sup>70</sup>

Khuri et al make the same point. NSQIP uses two very basic outcome indicators: mortality and morbidity. In that sense, it is not vastly different from the measuring currently done in the NHS, which shows whether a patient has died or been readmitted (both of which account for a minority of cases) but not the extent to which their health has actually improved. "There are other dimensions of surgical outcome that could be incorporated into the NSQIP to more thoroughly assess the quality of surgical care", the authors stress. "The most important of these are long-term survival, functional outcomes, quality of life, and patients' satisfaction."<sup>71</sup>

## Healthcare Commission Star Ratings

### Summary

- Where information does not credibly relate to quality, its impact on both providers and patients is degraded even if it is easy to understand.
- Star ratings were linked to the 'Kudos and Censure' model by financial incentives and by the opportunity to apply for Foundation Trust status.
- When compared to performance in Wales, star ratings did improve performance in England.

### Purpose

The manifest rationale for publishing star ratings was that they gave a rounded assessment of performance of NHS organisations for the public. The latent purpose was to put pressure on NHS chief executives to deliver on the Government's priorities. Star ratings amalgamated the scores from a variety of different targets to produce a single summary score that could be mapped into one of four rankings from zero to three stars. In this way ministers defined "failure" (zero rated) and "success" (three stars).<sup>72</sup>

### Selection

Two sets of indicators were used in star ratings of acute hospital trusts from 2001 to 2005.<sup>73</sup> The first comprised nine key targets, of which six refer to waiting times. The other three were achieving a financial balance, hospital cleanliness, and improving the working lives of staff. The second batch of indicators comprised about 40 targets organised into a "balanced scorecard". These reflected ministerial priorities, included a subset of the large number of targets in the priorities and planning framework for 2003-2006, and satisfied the technical criteria of being applicable nationally, measurable, capable of being captured by indicators, and stable over time.

70. Jha et al, 2003

71. Khuri et al, 2002

72. Bevan G, "Setting Targets for Healthcare Performance: lessons from a case study of the English NHS". National Institute Economic Review, 2006 and Bevan G, Hood C, "What's Measured is What Matters: Targets and Gaming in the English Public Healthcare System", Public Administration, 2006 and Bevan G, Hood C, "Have targets improved performance in the English NHS?" British Medical Journal, 2006

73. From 2001 to 2003, star ratings also included assessments from clinical governance reviews by the CHI on the seven technical components of clinical governance: public, service user, carer and patient involvement; risk management; clinical audit; clinical effectiveness; staffing and staff management; education, training and continuing personal and professional development; and use of information.

### Methods

The principal methodological innovation of star ratings was the combining of different kinds of information on performance to produce a summary score. The methods used were criticised for being arcane<sup>74</sup> in requiring levels of measurement that the data did not justify.<sup>75</sup> The scoring system was designed to use each year's data to register improvements in performance (for example, more three-star hospital trusts) and to avoid volatility in the star ratings of individual trusts. This resulted in what appeared to be arbitrary selections of thresholds in measuring performance against key targets.<sup>76</sup>

### Presentation

The public presentation of the star ratings was designed to make an impact by giving a single summary score and by reporting targets in a way that was easy to understand:<sup>77</sup>

- “key targets” were reported symbolically: ✓ achieved, – underachieved, ✕ significantly underachieved;
- each focus area of the “balanced scorecard” was reported as high, low or medium;
- each indicator within each focus area was given a ranking: 5 significantly above average, 4 above average, 3 average, 2 below average, 1 significantly below average.

### Dissemination

Star ratings were published on the internet and as paper reports. They were also covered by national daily newspapers, local newspapers, and national and local television. Two professional journals also reported the results (the *British Medical Journal* for physicians and the *Health Service Journal* for managers). Star ratings were published within months of the end of the financial year on which they were based and before Parliament went into summer recess.

### Model of improvement

Star ratings explicitly gave trusts incentives through a range of rewards and sanctions, such as an extra £1m to high performing trusts and the possibility of applying for foundation hospital status. However, reviews of star ratings also identified that “kudos and censure” played a very important part in trusts’ responses to them.<sup>78</sup>

### Impact

There is evidence of two kinds of impacts of star ratings: reported improvements against ‘key targets’; and evidence of gaming,<sup>79</sup> which is known to have been endemic when targets were used in centrally planned economies<sup>80</sup> and in the public sector.<sup>81</sup> Improvements on three ‘key targets’ for three different types of hospital waiting times when compared with Wales<sup>82</sup> (where targets were not linked to sanctions and rewards), were as follows:

- *Time spent waiting in A&E.* The key target in England from January 2005 for accident and emergency departments was that 98 per cent of patients were to be seen within four hours, which was achieved. The comparable performance in Wales was 89.4 per cent and 91.9 per cent in June and September 2005. The National Audit Office examined performance in accident and emergency in England, in 2002,<sup>83</sup> and reported that improved performance and increased patient satisfaction was achieved despite increasing use of emergency services.
- *First elective hospital admission.* The key target for the maximum wait was that no one should be waiting more than nine months by the end of March in 2004 (and 2005). This was achieved in England in each of those years. In Wales the percentages of patients who did so were 22 and ten per cent in March 2004 and in March 2005.

74. Klein R, “Mr Milburn’s good hospital guide”, *British Medical Journal*, 2002

75. One problem being the derivation of mean values across indicators at the ordinal level of measurement.

76. In the star ratings for 2003, for example, the methods of scoring ‘key targets’ implied that for patients waiting longer than 12 months, going from 10 to 11 was twice as bad as going from two to three.

77. Details were available in supporting technical papers.

78. Mannion R, Davies H, Marshall M, “Impact of star performance ratings in English acute hospital trusts”, 2005

79. Bevan, G and Hood, C, Have targets improved performance in the English NHS?, *British Medical Journal*, 2006 and Bevan, G, and Hood, C, What’s Measured is What Matters: Targets and Gaming in the English Public Health Care System, 2006 and Bevan, G., 2006, Setting Targets for Health Care Performance: lessons from a case study of the English NHS. National Institute Economic Review, 2006

80. Berliner, J, Soviet industry from Stalin to Gorbachev, 1998 and Kornai, J., Overcentralisation in economic administration, Oxford University Press, 1994.

81. Smith, P, On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 1995

82. Auditor General for Wales, 2005, NHS waiting times in Wales. Volumes 1 and 2, Cardiff, The Stationery Office, 2005 and Auditor General for Wales, 2006, NHS Waiting Times: follow-up report, Cardiff, The Stationery Office, 2006

83. National Audit Office, 2004, Improving Emergency Care in England, London: The Stationery Office, 2004 [http://www.nao.org.uk/publications/nao\\_reports/03-04/03041075.pdf](http://www.nao.org.uk/publications/nao_reports/03-04/03041075.pdf) (last accessed 03/05/07)



- *First outpatient attendance.*<sup>84</sup> The key target for the maximum wait was that no one should be waiting more than 17 weeks by the end of March 2004 and throughout 2004/05. In 2005, Wales, with a population only six per cent of England, had nearly four times as many outpatients waiting more than three months.

However, these successes in delivering political priorities were countered by concerns that clinical priorities were distorted, that gaming was encouraged, and that many aspects of healthcare, not covered explicitly by star ratings, were ignored. In 2004, The Healthcare Commission chairman, Sir Ian Kennedy, stated that the Commission would “move...away from ratings only about specific targets to painting a much richer picture, having to do with performance across a number of domains”.<sup>85</sup> A Consumers’ Association survey earlier that year found that “less than half of those surveyed were aware of star ratings and almost half of these were unlikely to use the ratings to help them choose a health service”.<sup>86</sup> Arguably, the ratings did not reflect the quality of care (placing an equal value on cleanliness and mortality rates) but rather show the effectiveness of the trust’s management team.

## The London Patient Choice Project

### Summary

- The LPCP is very different from the other four case studies, as the information that it produced was published primarily to allow patients to choose providers.
- The LPCP demonstrated that although patients wanted and made use of information on quality and performance, it was no coincidence that they also liked reduced waiting times for their operations, free transport, and support in making choices. Additionally, the number of hospitals per square mile in the London

area is unique in England and Wales. It is not clear that the LPCP results would be replicable elsewhere in the country. In addition the simple comparison of waiting times may not be analogous to the more complex judgements of quality of care.

### Purpose

The LPCP was established to improve choices for patients who were clinically eligible for treatment and who had been waiting for treatment at an NHS London hospital beyond a target waiting time (initially six months, though this was later reduced). As the end of the target waiting time approached, patients were given an opportunity to choose from a range of alternative providers who had the capacity to offer earlier treatment.<sup>87</sup> The project had four overall objectives:<sup>88</sup>

- to develop the necessary capacity to treat patients expected to exercise choice;
- to develop a working patient choice system;
- to learn how to improve the design of the system and feed lessons into future London and national programmes;
- to improve patient waiting times and patient satisfaction.

The LPCP was intended to change the perception of the NHS as a system that limited patient choice, as Coulter et al point out.<sup>89</sup> A study in eight European countries, including the UK, had found strong support for the notion of free choice of provider: 92 per cent for primary care doctors, 85 per cent for specialists, and 86 per cent for hospitals.<sup>90</sup> However, British people were among the most dissatisfied with the actual opportunities for exercising choice. Only 30 per cent said that these were “good” or “very good”, compared to 73 per cent in Spain and 70 per cent in Switzerland. This dissatisfaction was con-

84. Details of ratings: <http://www.performance.doh.gov.uk/performance/performanceratings/index.htm> (last accessed 03.05.07)

85. Hansard: Select Committee on public administration 29 January 2004. <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmpublicadm/41/4012910.htm> (last accessed 03.05.07)

86. Q&A: NHS star ratings 2005, Guardian Unlimited, <http://society.guardian.co.uk/health/story/0,,1536445,00.html> (last accessed 03.05.07)

87. Burge P, Devlin N, et al, “London Patient Choice Project Evaluation”, RAND Corporation: Santa Monica, 2005

88. Dawson D, Jacobs R, Martin S, Smith P, “Evaluation of the London Patient Choice Project: System Wide Impacts. Final Report”, Centre for Health Economics, University of York, 2004

89. Coulter A, et al, “Patients’ Experience of Choosing Where to Undergo Surgical Treatment: Evaluation of London Patient Choice Scheme”, Picker Institute Europe, Oxford, 2005

90. Coulter A, Magee H, “The European Patient of the Future”, Open University Press, 2003

firmed by a MORI poll. More than a third (38 per cent) of those interviewed said that NHS patients did not have any choices, and a further third said they didn't know what choices they could have. Two thirds (66 per cent) said that they felt NHS patients should be able to choose where to have their operations.<sup>91</sup>

### *Selection*

The project was extended from cataract surgery to include patients awaiting specific procedures in ear, nose and throat surgery (ENT), general surgery, ophthalmology, orthopaedics, and urology. Gynaecology and plastic surgery were introduced on a pilot basis in southeast London only. The selected conditions or procedures in each specialty were those for which patients were most likely to have to wait for more than six months. For each procedure an agreed patient care pathway (clinical plan) was developed and agreed between the trusts and the LPCP team. Patients were excluded from the scheme if their "originating trust" (OT) could guarantee them treatment within eight months of going on the waiting list (later reduced to seven months), or if there were agreed clinical reasons why treatment by an alternative "receiving trust" (RT) was inappropriate. Clinical exclusion criteria were specified by the LPCP team after wide consultation with clinicians in the relevant specialties in both originating and receiving trusts and after reviewing the literature to determine best practice.<sup>92</sup>

### *Methods*

The LPCP was co-ordinated by a central team. If patients on the in-patient waiting list of a London trust were to be offered the choice of an alternative provider, the OT had to agree to co-operate with the scheme and was linked up with two RTs, which could be NHS or private, or new treatment centres with spare capacity. An eligible patient was then offered the choice of

remaining with the OT or obtaining more rapid treatment at either of the two alternatives.

### *Presentation and Dissemination*

Independent patient care advisers (PCAs) were responsible for liaising with patients throughout the process. Once placed on the waiting list for surgery, patients attending outpatient clinics at OTs were to be informed in general terms about the choice scheme. Those patients whose medical conditions were too complex to be considered for a transfer to an RT were to be identified and screened out of the LPCP system. Staff at OTs were responsible for validating the waiting lists and sending names of eligible patients to the advisers. Validation was supposed to take place when patients had been on the waiting list for about four months. If patients were considered ineligible for the scheme at this stage, staff were required to inform them of the reasons for this.

Eligible patients were sent a letter giving them advance warning that a PCA would telephone them to discuss their options and an information booklet outlining the scheme in more detail. When patients reached a specific date on the waiting list, an adviser would offer them the option of going to an alternative hospital and would answer their questions.

If the patient accepted the offer of an alternative, the adviser was responsible for booking an "operative appointment" – a combination of a clinical validation, an outpatient appointment and a pre-operative assessment. The adviser would offer continued telephone contact with the patient, and support if any problems occurred. The PCA was also responsible for keeping the patient's GP informed if the operation had been scheduled to take place at an RT. Finally, a date would be specified for the operation. The intention was that the treatment would be completed within eight-and-a-half months of joining the waiting list.<sup>93</sup>

91. Patient Choice, MORI study conducted for BUPA Health Debate, 2003

92. Coulter A, et al, "Patients' Experience of Choosing Where to Undergo Surgical Treatment: Evaluation of London Patient Choice Scheme", Picker Institute Europe, Oxford, 2005

*Model of improvement*

As the title of the project suggests, the expectation was that patients would choose providers who were able to offer earlier treatment. Information was essential to enable them to make this choice.

*Impact*

Coulter et al (2005) reported on a study of patients from five OTs, who were sent postal questionnaires before they had been offered a choice of hospital and after they had been discharged from hospital. There were also in-depth interviews with sub-samples in each group. They found that:

- less than a third (32 per cent) of patients apparently eligible for the scheme were actually offered a choice of hospital. Of those who were offered the opportunity to go to an alternative hospital, two thirds (67 per cent) chose to do so;<sup>94</sup>
- those who were in more pain and who felt that their home hospital had a poor, or only fair, reputation were significantly more likely to choose to undergo treatment elsewhere;
- most patients who opted for an alternative hospital were treated in NHS treatment centres (82 per cent) and were more positive about their experience in those centres (or private hospitals) than in an ordinary NHS surgical department. Patients treated at alternative

hospitals were significantly more satisfied with their hospital experience than those treated at their home hospital;

- patients valued the provision of free transport and the support they received from PCAs in guiding them through the process. This included helping them to make a decision, and co-ordinating arrangements between the hospitals if they had opted to go to an alternative hospital;<sup>95</sup>
- when deciding where to undergo treatment, patients tended to place greater emphasis on issues such as the location of the hospital, length of wait for the operation, travel arrangements and convenience for family and friends;
- one in three survey respondents expressed dissatisfaction with the amount of information received about the various hospitals: they wanted to know more about arrangements for follow-up care, quality of care, the qualifications and experience of surgeons, operation success rates, standards of hygiene and safety records;
- an overwhelming majority (97 per cent) of patients who had opted to go to an alternative hospital said that they would recommend the scheme to others.

The LPCP achieved its principal objective to provide faster access to good quality care.

93. In the event this procedure was not followed in each of the OTs exactly as specified. In some cases patients were not given the preliminary information about the scheme to avoid raising expectations among those subsequently deemed ineligible. Although the LCP team recommended early 'fitness for surgery' assessment, this was not common practice during the life of the project. The process of validating the waiting lists and determining eligibility also varied between trusts (Coulter et al 2005).

94. There was no evidence of inequalities in access to, or uptake of alternative hospitals by social class, educational attainment, income or ethnic group; but people in paid employment were more likely to opt for an alternative hospital than those not in paid employment.

95. Others went to other NHS hospitals in London (13 per cent) or private hospitals (5 per cent). Only one patient in the study received treatment at an overseas hospital.

---

# 4

## Patients, payment and providers

Our five case studies all operated in different ways, in different contexts and with different results. What can we learn from them? And can this learning be applied today in the NHS?

The Labour Government's NHS policies have gone through three dramatic shifts in emphasis.<sup>96</sup> In 1997, *The New NHS*<sup>97</sup> laid out policies that sought a "third way" between centralised direction and markets. The end of that search was marked in 2000 by *The NHS Plan*,<sup>98</sup> which introduced centralised direction in the form of targets that eventually became star ratings. *Delivering the NHS Plan*<sup>99</sup> in 2002 marked not continuity but another radical shift in direction, this time to an emphasis on markets and incentives. There are three key differences in the design of this market from that introduced in the early 1990s:

- an explicit recognition that the response to failure ought to be managed;
- hospitals are paid a fixed tariff by types of case, so they compete on other grounds (such as convenience, waiting times and quality);
- patient choice is used as a means of improving quality.

This raises two obvious questions. Do patients act on information about quality? Do providers respond to financial incentives?

Do patients act on information on quality?

Although, when asked, consumers say they want more information about the outcomes attained by providers,<sup>100</sup> a review by Marshall et al concluded that "most experts do not believe that consumer pressure will be an important mechanism to stimulate quality improvements for the foreseeable future".<sup>101</sup> A study into consumer use of the Pennsylvania report card system showed that, of 474 patients surveyed, only 56 were aware of the report cards at the time of their surgery and only a quarter of these said that the system had a significant impact on their choice of surgeon. Another more recent review observed that "when this information is published only a minority are aware of it, of those most do not understand it (including whether high or low rates of an indicator reflect good performance), trust it or use it (with problems with timely access and lack of genuine choice); and evaluations of later developments that addressed many of these potential barriers failed to demonstrate significant or sustained public interest."<sup>102</sup>

These reviews are an important corrective to the simplistic view that merely publishing information will result in patients switching providers. However, awareness of the availability of information and ease of interpretation both matter. In the Pennsylvania case, more than half of all respondents claimed that the report cards would have influenced their decision had

96. Stevens S, "Reform Strategies for the English NHS", 2004

97. *The New NHS: Modern, Dependable*, the Stationery Office, 1997

98. *The NHS Plan*, the Stationery Office, 2000

99. *Delivering the NHS Plan*, the Stationery Office, 2002

100. Edgman-Levitan S, and Cleary P, "What information do consumers want and need?", Health Affairs, 1996

101. Marshall et al, 2003

102. Bevan, G "Impacts of reporting healthcare performance", in press

they been aware of them, indicating that the problem was ignorance of, rather than lack of enthusiasm for the data.

Paying for performance – do providers respond to financial incentives?

Redesigning payment systems to pay doctors and institutions by results is increasingly fashionable. The new GP contract, backed by the QOF is internationally recognised as one of the most ambitious of such schemes.

Less discriminating funding mechanisms that pay for activity regardless of how necessary it is create a perverse incentive to over-treat the well-insured. As shown in the US, this drives up costs without adding value.<sup>103</sup> There is similar evidence to be found in Canada, Denmark and Scotland.<sup>104</sup> This problem has spurred attempts in the US to reward improvements in quality of care (rather than quantity) with extra payments:

- several health insurance plans in California award bonuses to physician groups for patient satisfaction and chronic care performance;
- an employers' programme called Bridges to Excellence offers doctors \$50 per patient for patient education and installing clinical information systems in their offices or surgeries;
- hospitals scoring highly on a range of measures (including heart attack, CABG, and hip and knee replacements) receive a percentage of their Medicare payments.

Critics fear also that, like central Government targets in the NHS, pay-for-performance schemes may distort clinical priorities and have other unintended consequences. The resistance of doctors themselves to the idea is another problem. A 2002 survey found that, while 80 per cent of doctors wanted clinics to receive more funding for quality improvement, only 38

per cent supported the idea of making direct payments to group or individual physicians in return for achieving such improvements.<sup>105</sup>

Literature assessing the success of financial incentive systems is very mixed, with some achieving desired changes, others proving ineffectual and still others creating perverse and unanticipated effects:<sup>106</sup>

- a literature review of 89 studies from a range of countries concluded that financial incentives had a positive impact on many performance variables, such as admission rates to hospitals, duration of hospital stays, and, crucially for our purposes, compliance with clinical practice guidelines;<sup>107</sup>
- several studies from the 1990s of financial incentives for screening and immunisation showed that these had little effect.<sup>108,109</sup> Yet a study of targeted financial incentives for influenza immunisation in a Medicare population showed that even a small-scale incentive could achieve a statistically significant improvement.<sup>110</sup>

It is unclear to what extent financial incentives add to quality improvement beyond what can be achieved through measurement and reporting alone. A comparison was made between the same measures of chronic illness management in California, where a financial incentive was offered, and the Pacific North West,<sup>111</sup> where feedback and reporting alone were used. In only one instance was there a significantly greater improvement in performance in California than in the Pacific North West. By contrast, a very recent study of the Premier/Medicare pay-for-performance scheme, which compared hospitals that merely reported information about their quality of care with hospitals that combined this with performance-related pay, found that the latter group registered "modestly greater improvements in quality."<sup>112</sup>

103. Porter M, Teisberg E, *Redefining Health Care: Creating Value-Based Competition on Results*, Harvard Business School Press, 2006

104. Gosden, T, Forland, F et al "Impact of Payment Method on Behaviour of Primary Care Physicians: A Systematic Review", *Journal of Health Services Research and Policy*, 2001

105. Rolnick et al, 2002

106. Doran T et al, "Pay-for-Performance Programs in Family Practices in the United Kingdom", *New England Journal of Medicine*, 2006

107. Chaix-Couturier C, Durand-Zaleski I, Jolly D, and Durieux P, "Effects of Financial Incentives on Medical Practice: Results from a Systematic Review of the Literature and Methodological Issues", *International Journal for Quality in Health Care*, 2000

108. Hillman et al, 1998

109. Ibid

110. Kouides et al, 1998

111. Rosenthal M, Frank R, Li Z, Epstein A. (2005) Early experience with pay for performance from concept to practice. *JAMA* 294, 1788-1793

There is also conflicting evidence about financial incentives in other public and private sector arenas:

- an evaluation of the impact of a performance-related pay system for teachers in England concluded that “the scheme did improve test score gains, on average by about half a grade per pupil.”<sup>113</sup> The caveats were that this did not apply to all subject teachers and that the researchers were unable to determine whether improvement represented extra effort or effort diverted from other professional activities;
- in executive compensation “direct evidence of the responsiveness of executive performance to financial incentives is minimal”;<sup>114</sup>
- pay-for-performance registered some improvements in the arena of in-job training but gave rise to gaming, such as providers targeting their efforts at clients for whom training provided the least added value, which reduced cost-effectiveness.<sup>115</sup>

In their wide-ranging review of evidence, Rosenthal and Frank concluded that “the current enthusiasm for pay for performance in healthcare rests more on conceptual than empirical foundations”. However, their research suffered from at least three flaws. First, it was not clear “that the findings from the literature are indeed comparable to the broader efforts now envisioned, which would systematically identify and reward the best providers using multi-dimensional quality measures”. Second, many of the studies they examined involved very narrow sets of measures, such as compliance with guidelines on preventive healthcare. Third, they also suggest that the financial incentives offered in the studies may have been too small.

The mixed evidence on pay-for-performance is hardly surprising. If it were straightforward, value-based competition

would have been introduced already. Although the issue of financial incentives is no longer uncharted territory, simple solutions are unlikely to be adequate given the complexity of healthcare.<sup>116</sup>

We now consider a controlled experiment that illuminates the quandaries about patient choice, financial incentives, and publication of information.

### Does publishing add any value?

The State of Wisconsin has often been a venue for innovation in public services. Advocates of school vouchers continue to cite the famous piloting scheme in Milwaukee in the 1990s. Equally significant, if less well known, are its pioneering efforts in healthcare. This case study will examine the efforts of a major purchasing group in the state to measure quality of care, assessing its results and determining the lessons to be drawn for the NHS.

The Employer Healthcare Alliance Cooperative is a not-for-profit co-operative of employer purchasers of healthcare. Set up in 1990 by seven local employers, it is based in Madison, Wisconsin and has contracts with a range of hospitals in the area. Its membership now includes 157 employers and 73,000 individual employees and family members. What distinguishes the Alliance from many employer-purchasing groups is its determination to base its purchasing decisions on value rather than simply on cost. While many purchasers seek the cheapest care available, the Alliance tries to contract only with hospitals that provide care of a good quality.

As its website puts it: “We practise value-based purchasing. We’re committed to improving the quality of the healthcare system overall, not just getting discounts. Discounts alone aren’t the answer – managing the system to prevent over-use, under-use and misuse of care, and to reward the best performing provider is.”

112. Lindenauer P, “Public Reporting and Pay for Performance in Hospital Quality Improvement”, *New England Journal of Medicine*, 2007

113. Atkinson A, Burgess S, Croxson B, Gregg P, Propper C, Slater H, and Wilson D, “Evaluating the Impact of Performance related Pay for Teachers in England”, The Centre for Market and Public Organisation, University of Bristol, 2004

114. Rosenthal et al, 2005

115. Ibid

116. There are two other arguments that can be made to support this both of which relate to the incentivisation of specific process measures. First, Rodney Hayward has argued that most pay-for-performance schemes target ill-evidenced gold standards that require disproportionately greater efforts to achieve than the lower well-evidenced targets. This means that they have effectively become tools for advocacy groups for specific diseases to demand a bigger share of the healthcare cake. Consequently Pay for Performance in this manner achieves precisely the same effect as fee for service. Secondly, Porter and Tusberg are themselves profoundly critical of pay-for-performance schemes as not achieving the value-based competition that they advocate. They argue for real payment by outcome, but are unrealistically optimistic on the achievement of this.

The alliance endeavours to make the employees it serves “better healthcare consumers” through education, which earned it the 2001 national healthcare purchaser award. The most notable attempt it has made to drive up quality in local hospitals was a systematic reporting of their performance, QualityCounts, launched in 2001. QualityCounts consisted of two indices of adverse events (deaths and complications) across three areas of care (cardiac, obstetric, and hip and knee). It classified hospitals as “better than expected” (denoting fewer adverse events than expected), “as expected”, or “worse than expected”. The data was obtained from the Wisconsin bureau of health information.<sup>117</sup> Exceptional efforts were undertaken to make the report publicly accessible: it was mailed to employees covered by the alliance, included as a supplement in the local newspaper in Madison, made available online, and also distributed in hard copy to libraries and community groups.

#### *The impacts of information on providers*

There were 24 hospitals in south central Wisconsin that were in the Alliance service area and that received a “public report”. Hibbard compared them with the remaining 91 hospitals in Wisconsin, which were divided into two control groups of similar size, similar baseline levels of performance and similar characteristics such as average total inpatient days. One group received no report, the other a “private, confidential” report.

The study is unusual in that it examined the nature of changes in both outcomes and processes: asking hospitals how many of seven specified reforms to reduce complications in obstetric care they were implementing. An earlier study by the same authors found that in the nine months after the release of the public and private reports, “public report” hospitals had undertaken significantly greater quali-

ty improvement efforts than “private report” or “no report” hospitals. This study also revealed that hospitals expected the report to affect their public image but not their market share.<sup>118</sup>

The later study largely confirmed these findings. For obstetric care:

- In the *public report group* (whose information was given to them and was also published locally), a third of hospitals achieved statistically significant improvements in obstetric care in the two years following the report, and only 5 per cent showed a significant decline. Of the eight hospitals with baseline performances “worse than expected” only one was still at this level two years later. The average score of this group for implementing the seven specified reforms to reduce complications was 4.1 (and 5.7 for those which showed an improvement two years later);
- In the *private report group* (whose information was given to them but not published locally), 25 per cent of hospitals improved, and 14 per cent declined. Two thirds of the hospitals with baseline performances that were “worse than expected” were still at this level two years later;
- In the *no report group* (whose information was neither given to them nor published locally), about 12 per cent improved. Nearly two thirds of the hospitals with baseline performances that were “worse than expected” were still at this level two years later.

Results for cardiac care echoed those for obstetric care, but did not reach statistical significance, which the authors explained by noting that there were far fewer hospitals with poor scores in cardiac care. There was not enough baseline variation in hip and knee care for the authors to study it for post-report changes.

117. Hibbard J, Stockard J and Tusler M, “Does publicizing hospital performance stimulate quality improvement efforts?”, *Health Affairs*, 2003

118. *Ibid*

These results demonstrate the additional power of public reporting to change the behaviour of providers. However, its effects on local patients are interesting and at first sight contradictory. The QualityCounts report was well known both among employees covered by the alliance and by the general public. Of employees covered by the alliance, 57 per cent had seen the report, and 61 per cent had been exposed to it to some degree, such as by seeing the report, reading about it in the press or hearing about it from another person. Immediately after the report was released, 39 per cent of the general public had been exposed to the report. Two years later this fell to 24 per cent.<sup>119</sup>

In addition, those who had seen the report gained a good understanding of the quality of local health services. Immediately after the report was published about 35 per cent of respondents correctly identified the high-performing hospitals, and 63 per cent the poorly-performing hospitals. Of those that had not seen the report, the proportions were 21 per cent and 39 per cent respectively. Some two years later there was little change in identification of the high-performers, but a decline in identification of the low-performers.

There was, however, no significant change in market shares during the period from just before release of the report to two years after. There were no shifts away from low-rated hospitals or toward higher-rated hospitals in overall discharges, or in obstetric or cardiac care cases during any of the examined post-report time periods.<sup>120</sup>

This raises interesting questions about the motivations of both patients and providers. Given robust and easily understood data about the quality of care, patients did not use it to change providers, echoing the results of our survey and the experience of those investigating cardiac surgery in New York. Early on, local hospitals expected their report to affect their

public image but not their market share, and they were correct in this expectation. The main driver of change seems to have been a desire to maintain a good reputation, rather than fear of an actual decline in revenue.

That said, all hospitals were “slightly negative” about the general idea of publicly reported outcome measures. When asked to rate the validity of these measures from 1 (denoting not valid at all) to 5 (denoting very valid), the “public report” group was the most sceptical (mean rating of 2.1), the “private report” group was the most positive (mean rating of 2.6), and the “no report” group was somewhere in between. There were similar results when the groups were asked how appropriate the QualityCounts report was for use by consumers, and how effective it was as a way of driving up quality of care. Within the “public report” hospitals, those with the lowest scores were the most sceptical. Nevertheless, the most impressive improvements were achieved by precisely those hospitals that were most hostile to the report, the poorly-performing “public report” group.

The conclusion to be drawn from the QualityCounts experiment is that public reporting of outcomes data is an optimal way of securing gains in the quality of care. However, as Hibbard et al observe, there are three conditions that must be met if such reports are to enjoy similar success elsewhere. The report must be widely distributed throughout the target community, the information itself must be presented in a way that is easy to understand and the hospitals themselves must know that another public report will be published in the near future so that they have an incentive to address any shortcomings revealed in the first report.

All of these seem to be common sense, and the first in particular would be relatively simple to achieve in Britain. The methods used in Wisconsin to ensure a wide distribution of the report – local

119. Ibid

120. However, the failure of the report to impact on market share may be at least partly due to the particular circumstances of Wisconsin. An unusual feature of the state's health market is the close alignment between hospitals and physician groups. Physicians tend to practise only at hospitals aligned with their own physician hospital organisation (PHO). Over 85 per cent of physicians in the area served by the alliance operate in such a system. Whereas only 30 per cent of hospitals nationally have either formal or informal alignments with PHOs, almost all hospitals in south central Wisconsin do. The significance of this for patients is that if they decide to change hospitals they are likely also to have to change physicians, something many patients are understandably reluctant to do. The implication is clear: “that in markets where this high degree of alignment is not present, a public report could raise concerns about both reputation and market share, motivating improvements through both of these pathways.”



newspapers, direct mailing, websites and the provision of hard copies in libraries and other public facilities – can also be employed in this country. Indeed, the fact that the primary care sector is so much more widely and frequently accessed by patients in Britain than by their American counterparts is an advantage in disseminating information. GP surgeries, already highly visible features of most communities in this country, can become the main point of call for information about the quality of local health-care providers.

The second condition is not without hazards. Making a report easy to evaluate may sometimes mean simplifying the data to a degree that would meet with objections from providers. League tables, in particular, are likely to be criticised as crudely and baselessly presenting some hospitals as simply “better” than others, when the reality may be much more complicated and ambiguous. League tables tend to emphasise differences in rank which may have no practical difference in quality, thus creating false criteria for choices. In Wisconsin’s case, going to the first or fifth ranked hospital would have made very little difference to the patient. A league table approach is likely to increase gaming and arguing about numbers, as the effect of very small changes could make a big difference in the ranking.

Wisconsin succeeded in creating data that was easy to evaluate because it eschewed league tables in favour of measuring against expected standards. Anecdotal discussions about unpublished polling data in the United States suggest that patients understand this, with three quarters of citizens wanting comparison against expected performance rather than between physicians.

### Conclusions

Before going on to consider practical examples of how to use information about

quality as a more effective tool let us recapitulate the key lessons from our review of experience to date, our survey of patients and our review of currently available data.

- Measuring quality of care and publishing the resulting data makes health services more accountable to patients and the public. It may also lead to improvement in these services by encouraging patients, providers or both to behave in different ways.
- When it is done well, as in the case of the VA, or the QualityCounts initiative in Wisconsin, measurement and publication successfully improves the quality of care.
- It is, however, hard to do well. There are two distinct difficulties: the technical challenge of measuring quality and presenting the information in a way that improves quality.
- The effects of linking formal incentives to quality measurement such as pay for performance are ambiguous. Such incentives encourage providers to take the quality agenda seriously, but also encourage unintended perverse responses such as gaming and falsification of data. It is not clear that the extra costs of providing a financial incentive encourage sufficiently better performance.
- Publication is important. As the QualityCounts initiative shows, publishing quality measurements, rather than just reporting confidentially back to providers, seems to increase the likelihood of changing their behaviour.
- However, evidence from the US and the UK is at best ambiguous about whether patients want this information to enable them to choose between different providers. Our own survey confirms unpublished results from the US that suggest that this is not a priority. In addition, very well studied publication schemes in the US, such as

QualityCounts and cardiac surgery mortality in New York state showed that publishing data did not lead to changes in hospital market share. The LPCP by contrast, appears to have been successful, although this may reflect both the large infrastructure put in place to support choice, and the relatively simple metrics needed to understand how long one waits (a somewhat simpler proposition than comparing the quality of two services).

- There are currently many gaps in available data, both in the UK and globally.

This is partly due to a lack of sub-hospital information (i.e. information at the level of an individual service), but there is also a lack of outcome data, other than measures of failure. There is not an internationally accepted list of performance measures that the UK can just pick up and use.

In the following chapters we consider some of the technical problems of measurement and then suggest ways of filling some of the gaps in the information that is currently available.

---

# 5

## Technical difficulties

### Introduction

The previous chapters have sought to demonstrate the results that have been achieved by efforts to measure and improve the quality of healthcare. The case studies of the VA, Wisconsin and New York show that these measurement systems, when combined with active programmes to drive up the standards of poorly performing providers, have led to dramatic improvements in the quality of care received by patients.

Given these examples of apparently successful use of measurement, one may wonder why the practice is not more widespread. At least part of the answer is that it is hard to perform. There are difficulties in establishing measurement regimes that stimulate improvements without also producing unintended consequences. These include the design of the incentive system around the measurement system; ensuring adequate data quality; and designing analysis that is sufficiently sophisticated to allow for the complexities of healthcare, and is easily understood by its target audience. There are also problems in deciding where and how to publish the results. These difficulties are technical rather than ideological.

No credible attempt to argue the case for measurement can fail to consider the obstacles faced by such an approach. This chapter examines three of the most significant: the question of whether it is better to measure outcomes or processes; the distorted clinical priorities that can result from the rigid application of certain measures; and the fundamental problem of whether what is intended to be measured is what is actually being measured.

What to measure: outcomes or process?

Most health experts caution against using outcomes to provide a judgement of the quality of care. For them, outcomes are determined by a range of variables, of which quality of care is only one. The most obvious of these is case-mix, and the complexity and intensity of the care that they receive as a result. “Hard cases”, those patients with poor health and/or suffering from complications, are less likely to attain good outcomes than healthier patients receiving the same quality of treatment. Of course, almost all outcome measures go to considerable lengths to adjust for case-mix, but, as Lilford warns, this “can lead to the erroneous conclusion that an unbiased comparison between providers then follows...Making judgements about quality of care on the basis of risk-adjusted comparisons cannot guarantee that like is being compared with like”.<sup>121</sup> This is the “case-mix fallacy”.

However, the variables affecting health outcomes do not stop at the case-mix of patients. There is, for example, a problem of definition: whether or not a live birth or fractured neck of femur has taken place is self-evident, but what counts as a heart failure or infertility or myocardial infarction?<sup>122</sup> Much depends on the discretion and the subjective judgement of the clinicians on the scene. The variation in definition can affect recorded health outcome data as much, if not more than, variations in the quality of care.

These problems are not merely theoretical. Apparent variations in outcome may be

---

121. Lilford et al, "Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma", 2004

122. Lilford, et al. 2004

so influenced by case-mix variability and definition that they do not relate closely enough to quality of care to be cited as a measure of how well a hospital performed. Several studies have shown examples of poor correlation between quality of care and outcome in practice. One review concluded that most hospitals in the top 5 per cent for mortality rates (the “worst” hospitals) will not be in the lowest 5 per cent for quality of care. Three separate studies found no association between outcome and quality of care in acute medical care,<sup>124</sup> myocardial infarction<sup>125</sup> and congestive heart failure or pneumonia.<sup>126</sup> Other work has discovered relatively weak associations for stroke,<sup>127</sup> hip fracture<sup>128</sup> and myocardial infarction.<sup>129</sup>

This problem is exacerbated when a league table ranking is used to judge the relative performance of different hospitals, as the different rankings of hospitals on the table may not reflect any meaningful differences between their performances. Lilford et al’s conclusion, typical of the views of many health policy specialists and clinicians, is that although subjecting healthcare to some kind of performance measurement is valuable, “league tables of outcomes are not a valid instrument for day-to-day performance management by external agencies.”<sup>130</sup>

This prompts the question of what is an appropriate way of measuring the quality of care. The response offered by Lilford et al is that, although outcomes should still be measured and used by hospitals themselves for their own performance management and quality improvement, measuring adherence to clinical process has more value.<sup>131</sup> In other words, clinical care should be judged according to whether appropriate medical interventions are made during a patient’s care, such as providing beta blockers after acute myocardial infarction, using lower tidal volume in acute respiratory distress syndrome, or even just providing influenza vaccinations to those at risk. Process measures are already used in both the US and the UK. The American Medicare system and our own GP contract

include measures such as whether the correct antibiotics are prescribed for pneumonia, and the influenza vaccination rate.<sup>132</sup>

“ Clinical care should be judged according to whether appropriate medical interventions are made during a patient’s care ”

Process measures can claim two major advantages over outcome measures. The first is clarity: failing to meet agreed and evidence-based standards of care, such as providing beta blockers, can be regarded as actual bad care, rather than just an indicator of bad care. Unlike process data, a high mortality rate does not tell a hospital precisely what the cause of the problem is. Lilford et al ask us to consider three of the most infamous cases of clinical failure in the NHS in recent years: the Bristol Royal Infirmary, Dr Harold Shipman, and a cardiac surgeon with exceptionally high mortality rates. These were, respectively, the products of poor systems of care and scrutiny, a murderous personality, and a surgeon operating while suffering from an undiagnosed brain tumour. “Even in these most extreme situations”, note the authors, “we are unable to reliably use outcome data to judge where the quality of care was deficient.”<sup>133</sup>

Secondly, there is efficiency, as the measurement can be taken at or near the moment care is delivered rather than waiting for 30 days (or more, depending on the measure) to assess the health outcome. The information costs attached to process measures are generally lower. One study estimated that variations in quality of care may cause a 10 per cent difference in mortality across hospitals and that, although 3,619 patients from each hospital would have to be assessed to detect this, it would take only 48 from each hospital to detect the corresponding difference in process measures.<sup>134</sup>

123. Thomas J and Hofer T, “Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care”, 1998

124. Best W and Cowper D, “The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients”, *Medical Care*, 1994

125. Jenks S et al, “Interpreting hospital mortality data: the role of clinical risk adjustment”, *JAMA*, 1988

126. Park, R et al, “Explaining variations in hospital death rates: randomness, severity of illness, of care”, *JAMA*, 1990

127. McNaughton H et al, “Relationship between process and outcome in stroke care”, *Stroke*, 2003

128. Freeman C et al, “Quality improvement for patients with hip fracture: experience from a multi-site audit”, 2002

129. Krumholz H et al, “Evaluation of a consumer-oriented internet healthcare report card: the risk of quality ratings based on mortality data”, *JAMA*, 2002

130. Lilford et al, 2004

131. Ibid

132. Ibid

133. Ibid

134. Mant J and Hicks N, “Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction”, *British Medical Journal*, 1995

Does this mean that we should abandon publication of outcome measures in favour of process measures? Arguments against this view cite the fundamental purpose of health-care (it is ultimately about improvement of health and alleviation of suffering, not adherence to protocols), and some of the weaknesses of process measures. Process measures may restrict innovation by requiring providers of care to perform certain clinical interventions, whereas outcome measures essentially leave it up to them to decide how to achieve good outcomes. We consider this in greater detail in the next section. Process measures may also be prone to being unimaginatively followed in ways that create unintended or perverse consequences. As Smith and others observe, “many measures of process are highly vulnerable to misrepresentation by care providers, and may induce seriously dysfunctional behaviour such as gaming.”<sup>135</sup>

“ Ultimately, there seems little reason to make a choice between process and outcome measures, as they can complement each other ”

Ultimately, there seems little reason to make a choice between process and outcome measures, as they can complement each other. The idea that different types of measures serve as “tin openers and dials” is apposite here. Outcome measures, as long as they are designed and applied with sophistication, can act as tin openers, exposing the existence of a problem that needs to be investigated further. Process measures, if they can be credibly linked to the outcomes, then identify whether there really is a problem and where exactly it is located. Among the advantages secured by combining both types of measures is that it may reassure providers that they are not being judged by incomplete and unhelpful data, but by well-rounded information

that identifies exactly what, if anything, they are doing wrong.

The problem of rigidity: measures are not the same as guidelines

Another criticism of process measures is that they are an overly rigid and inflexible version of practice guidelines. They can encourage clinicians to make interventions in order to “tick a box” rather than because they judge them necessary for the health of the patient.

Walter et al examined the particular case of the San Francisco VA Medical Center, where clinicians were told that failing to raise the centre’s rate of colorectal cancer screening from 58 per cent to the VA national target rate of 65 per cent could incur financial penalties. The implication was that low screening rates indicated poor quality of care, when in fact in some cases it meant the opposite. For many patients, such as those already suffering from severe conditions or with strong objections to screening, more harm than good may be done by subjecting them to screening. In such cases, good quality of care will take the form of doing the very opposite of what the San Francisco doctors were instructed, and given incentives, to do.

The authors identified that the problem lay in the blurring of the line between two forms of external monitoring that should be kept distinct: practice guidelines and process measures. The VA’s 65 per cent target was derived from practice guidelines, which by themselves are useful and evidence-based. (Trials show that patients who have been screened do generally suffer less colorectal cancer mortality.) However, as the authors stressed, “performance measures are not the same as practice guidelines.”<sup>136</sup>

Guidelines are to be applied with discretion by clinicians themselves. Inherent in their name is the acceptance of a grey area in many clinical decisions: interventions

135. Smith P, “Some principles of performance measurement and performance improvement”, Report prepared for the CHI, 2002

136. Walter et al, “Pitfalls of converting practice guidelines into quality measures”, JAMA, 2004

that are generally desirable may be inappropriate in particular cases. Process measures, by contrast, set standards that indicate failure if not met. They specify what are “good” and “bad” medical care, often assigning rewards and penalties respectively. In short, performance guidelines are advisory, while process measures are mandatory.

As with the criticism of outcome measures made by Lilford et al (that outcomes do not serve as accurate signifiers of quality of care), this problem is not just theoretical. The Walter study of the San Francisco case uncovered the practical results of the perverse incentives produced by rigid process measures. Although patients with severe conditions associated with life expectancies of only five to ten years are unlikely to benefit from screening (and may be harmed by it), a significant percentage of the 229 patients audited for colorectal cancer screening in at the San Francisco centre in 2002 had these characteristics.<sup>137</sup> Thirty-five per cent were 75 years old or more, while 24 per cent had severe disease. These included a 94-year-old man with metastatic prostate cancer, an 89-year-old woman with severe dementia, and a 76-year-old man with end-stage renal disease who died two months before the screening date.

However, as the authors maintain, these problems are not proof that process measures are inherently flawed but that they should be carefully designed. In particular, measures should only apply to the care of patients for whom evidence shows that the interventions in question would do more good than harm.<sup>138</sup> Process measures are often scored as the number of patients who receive an intervention (the numerator), divided by the number of patients who were eligible for it (the denominator). The problem with the VA measure is that the denominator included patients who did not want to be screened or for whom screening would not have been beneficial. The fact that they were not in the numer-

ator counted as a failure. Allowing the denominator to determine patient preferences and clinical judgements would only require doctors to document systematically the discussions they have with patients and the recommendations they make to them.

Construct validity: do the same measures give different results?

The most important requirement for an indicator of performance is “construct validity”; the fact that it actually measures what it intends to measure. If two different measures purporting to measure the same thing produce different results, at least one of them can be said to lack construct validity.

Little work has been done on examining the construct validity of existing measures of healthcare, but some of it raises concerns. Brown and Lilford conducted a study of performance indicators in English primary care trusts (PCTs) to determine their construct validity. They compared pairs of indicators that they expected, or hypothesised, would measure the same thing to see if there was a correlation between the results that they produced. Their findings were not encouraging. Analysis of four indicators purporting to measure access to services (the QOF’s “access bonus”, the star ratings’ “access to quality services” category, Dr Foster’s “equity” rating, and the patient satisfaction survey’s “access and waiting” section) found “insufficient evidence to suggest that these indicators are measuring the same underlying concept...It is impossible to say whether any of the access measures are ‘better’ than the others.”<sup>139</sup>

The same was found of performance indicators that notionally measure different healthcare concepts, but which the authors hypothesise are sufficiently related for results to correlate strongly<sup>140</sup> They hypothesised, for instance, that PCTs with high star ratings or National Health Service Litigation

137. Ibid

138. Ibid

139. Brown C and Lilford R, “Cross sectional study of performance indicators for English Primary Care Trusts: testing construct validity and identifying explanatory variables”, *BMC Health Services Research*, 2006

140. Ibid

Authority (NHSLA) ratings would have low hospital mortality, since star ratings measure the overall quality of a primary care trust (including hospital care), while NHSLA measures safety procedures that can be reasonably considered to form a part of quality of care. Their analysis found, however, that trusts with high star ratings did not necessarily have low mortality rates. PCTs with no stars had a mean mortality ratio of 102.7, those with one star had 99.9, those with two stars had 100.2, and those with three stars had 101.5. Similar results were found for the NHSLA ratings.

Evidence of construct validity was found only for QOF and star ratings measures of screening and preventive healthcare. Their study, Brown and Lilford conclude, “casts doubt on whether any of the available performance indicators help the public to accurately assess the level of care received at their PCT.”

However the Brown and Lilford study was a very limited one, as they concede. The range of indicators they looked at was small and pertained only to those to which English PCTs are subjected. Moreover, there were some positive findings, such as the correlation to be found between indicators in screening and preventive care. We conclude that the Brown and Lilford work should not be seen as casting doubt on most performance measures in healthcare.

## Conclusion

This chapter has considered some of the more significant technical problems that need to be overcome in order to measure the quality of healthcare in a reliable way. Of themselves, these are not ideological objections to using measurement as an accountability or improvement strategy, but they do emphasise the need for technical excellence, clinical expertise and managerial skill in their implementation.

Many of the authors whose work has highlighted these technical difficulties continue to maintain that the basic principle of measuring quality of care is sound and should be maintained. Walter et al<sup>141</sup> are typical: “Despite the pitfalls that may occur in converting a practice guideline into a performance measure, the potential benefits of performance measures derived from evidence-based guidelines should not be ignored. The quality of medical care can be improved in many areas, and performance measures are tools that can help achieve selected goals.” Their call is for better designed measure, not for the abandonment or restriction of measures.

It is to answer this call – by identifying the best possible healthcare performance measures currently in existence – that much of this report is devoted.

141. Walter et al, “Pitfalls of converting practice guidelines into quality measures”, JAMA, 2004

# 6

## Patient Reported Outcome Measures (PROMs)

We have noted that the biggest gap in the reporting of quality in the NHS is in the area of outcomes of care where nothing disastrous happened. We know how many people died, but we know little about how successful care was for the majority who did not. The QOF has a few measures which judge how well a chronic condition, such as diabetes, is being managed. For the majority of hospital care, however, there is very little such measurement. This chapter suggests that asking patients directly how their everyday life is being affected gives data of real power.

It has been argued that data provided by patients offers an important adjunct to clinicians in the care of their patients.<sup>142</sup> Self-completed questionnaires with adequate measurement properties offer a quick way for patients to provide evidence of how they view their health – evidence that can complement existing clinical data. This information can be used to screen for health problems and to monitor the progress of these problems, once identified, as well as the outcomes of any treatment.

Patient-based outcome measures may also help change the culture of the NHS; an organisation which is far from universally patient-focused. It is a politically-led organisation, where the Government, not the patient, is the paymaster. Government targets have often had the perverse effect of taking focus away from patient-centred professionalism.<sup>143,144</sup> As the NHS Confederation wrote in 2006, “One way of ensuring that NHS organisations focus

on providing services that patients want is by including in their measures of performance a measure to encompass patient outcomes and experience”.<sup>145</sup>

The results of a 2005 Patient-Reported Health Instruments Group report provide encouraging signs that PROMs may be effective in positively improving patient involvement and ultimately might improve some important longer-term outcomes.<sup>146</sup> There are a number of relevant studies that have been completed since 2000:

- The London School of Hygiene and Tropical Medicine (LSHTM) carried out research for the Department of Health on the routine use of PROMs in treatment centres, between 2005 and the end of 2006. The aim was to identify which PROMs could best be used to measure accurately health changes in treatment centres for five specified surgical procedures which account for 15 per cent of NHS elective surgery.<sup>147</sup>
- The Orthopaedics and Trauma Unit at York Hospitals Trust has been collecting health outcomes data for patients receiving hip and knee replacements since March 2001, to determine how much patients’ health improved following surgery.
- BUPA has pioneered data collection of patient reported outcomes in hospitals. BUPA has reported particular value from using PROMs to generate ‘safety engineering charts’ which identify outliers from normal practice.

142. Tarlov AR et al, “The Medical Outcomes Study. An application of methods for monitoring the results of medical care”, *Journal of the American Medical Association*, 1989; Nelson EC, Berwick DM, “The measurement of health status in clinical practice”, *Medical Care*, 1989

143. In primary care, Government targets for prevention could generate a tension between doctor and patient: overweight patients, for example, could well feel stigmatised by Government policy and be less trustful of, and willing to seek help from, their GPs. At the secondary care level Government targets on the time patients waited to be seen in A&E, for example, could lead to a processing mentality by staff whose primary objective was to ensure targets were met, at the expense of patient-centred professionalism.

144. This has raised the fundamental question of the doctor’s role - was the doctor’s responsibility to do his or her best for the individual patient, or to help meet political targets even in local circumstances that might be incompatible with truly patient-centred professionalism?

145. Edwards, N, “Lost in translation: why are patients more satisfied with the NHS than the public?”, *NHS Confederation*, 2006

146. National Centre for Health Outcomes Development, *Patient Involvement and Collaboration in Shared Decision-making: A review. Report to the Department of Health*, 2005.

147. The procedures are Cataract surgery; Hip replacement; Knee replacement; Varicose vein procedures; Hernia repairs.



It is this final example that we examine in this section of the report.

Background

BUPA is the largest and best known private health insurance company in the UK. 3.1 million Britons have BUPA insurance, as well as another 2.9 million people internationally. It provides care through 35 acute hospitals and 245 care homes.

During the 1990s, BUPA became frustrated with the limits of existing measures of health outcomes used in the UK.<sup>148</sup> The NHS tended to do no more than measure various forms of failure, such as readmission and mortality, which account for a relatively small number of all patients, as Figures 1 and 2 show. The independent sector, with its more predictable elective

workload, had developed more useful process measures, which defined standard care pathways and monitored deviations from these pathways. However, neither the state nor the independent sectors were measuring what is most important: the extent to which patients themselves feel they have gained in health and well-being after receiving healthcare.

To fill this gap, BUPA began the search for a patient-reported outcome measure. As a journal article by a group of BUPA employees notes, “Only the patient understands their own health post-discharge with any degree of richness, so the use of a tool that assesses outcome from the patient perspective is essential.”<sup>149</sup>

SF-36

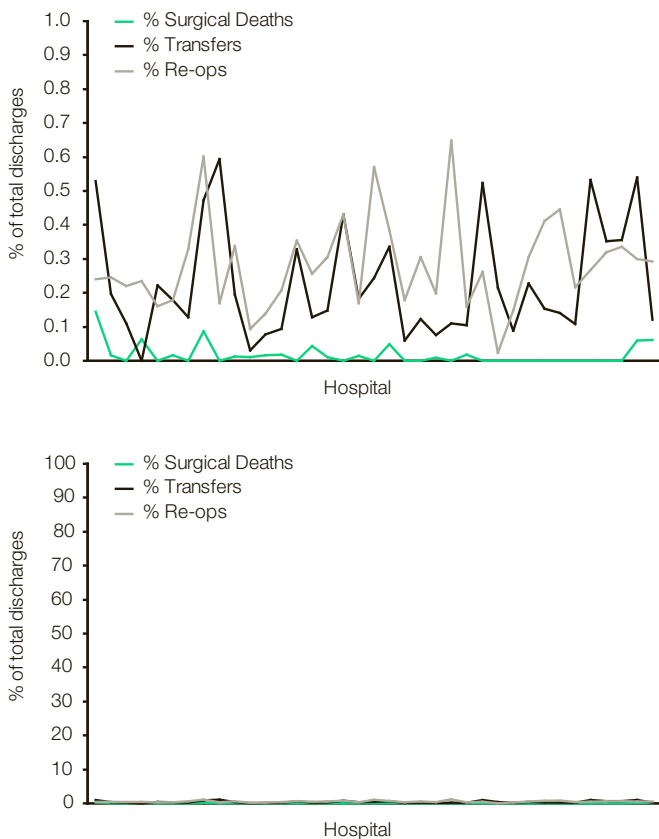
In 1998, BUPA eventually settled on the Short Form 36 (otherwise known as SF36)<sup>150</sup> as the measuring instrument it would use. SF-36, developed by the Medical Outcomes Study in the US, contains eight scores for dimensions of well-being, such as vitality and body pain, and two summary scores for physical and mental health. Unlike measures for specific conditions, such as the Oxford knee measure, it is a generic instrument suitable for the full range of medical interventions. It has been widely recognised as valid and reliable,<sup>151</sup> and its use in areas of care as diverse as orthopaedic surgery,<sup>152</sup> hysterectomy<sup>153</sup> and coronary artery bypass graft<sup>154</sup> documented.

BUPA asked patients to complete and return the SF-36 form pre-operatively (before admission) and post-operatively (three months later). 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.

The first form is collected by the hospital, while the second is mailed out to the patient with a freepost envelope. The

- 148. Vallance-Owen A et al, "Outcome monitoring to facilitate clinical governance; experience from a national programme in the independent sector", 2004
- 149. Ibid
- 150. 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
- 151. Ziebland S, "The short form 36 health status questionnaire: clues from the Oxford region's normative data about its usefulness in measuring health gain in population surveys", *Journal of Epidemiology and Community Health*, 2005
- 152. Nilsdotter A et al, "Comparative responsiveness of measures of pain and function after total hip replacement", *Arthritis & Rheumatism*, 2001

Figures 1 and 2: BUPA hospital rates for clinical indicators



change in scores from the first to second forms is the measure of health outcome. The results are fed back to hospitals every three months and to consultants every 12 months. Each provider can see their own results. The identity of the individual patients, however, is not revealed to hospitals or consultants. All of BUPA's hospitals participated in the scheme and, by 2004, over 100,000 patient cases of elective surgery had been recorded.

#### Problems and solutions

As Andrew Vallance-Owen, Medical Director at BUPA, noted that their original vision for outcome measurement proved to be a "little optimistic", problems began to emerge.

The single most significant problem was the question of how to present the data. Initially, league tables and histograms were used. However, these were criticised by providers for a range of reasons, including being difficult to interpret and providing only a snapshot view rather than presenting trends.

League tables were especially unpopular. As Vallance-Owen et al observe: "Outcomes are influenced by random variation as well as case mix, which can significantly alter a hospital's position in a league table. Nonetheless, there is an implicit assumption that hospitals located towards the bottom of a league table provide a worse service. This may be inaccurate and if so, is demoralising for staff".<sup>155</sup> It was felt that too many hospitals (about 1 in 20) were identified as outliers, a problem which meant that hospitals were wasting staff time on additional audits when their performance was actually within the normal range, while rewards were given to the "top" hospitals when their performance was actually due to chance.

To address these concerns, BUPA started presenting the SF-36 data in the form of Shewhart charts, otherwise known as con-

trol charts. Control charts plot data in relation to bands, or "audit lines", which are three standard deviations above and below the mean: any variation within these bands is attributed to common causes, while variations which exceed the bands are attributed to special causes.

There are several advantages to control charts. First, they are relatively easy to interpret. Second, looking into special causes is quicker and simpler than investigating common causes. It "enables one to examine clinical practice, to identify the factors behind apparently exceptional or potentially poor performance and hence facilitate clinical governance", which, after all, was the purpose of BUPA's efforts at measuring outcomes.<sup>156</sup> Third, they allow trends in performance to be identified, as opposed to just a snapshot view.

Another problem encountered by BUPA was the scepticism with which consultants initially responded to the measuring process. Many were hostile to the idea of being judged on the basis of the subjective opinion of patients, as opposed to more technical and quantitative measures of well-being, such as the range of movement of a joint in degrees or visual acuity as gauged by a Snellen chart.<sup>157</sup> From the spring of 2002, steps were taken to address these concerns. For example, following complaints from ophthalmologists, SF-36 was replaced by a specific questionnaire (the Visual Function 14) for phakoemulsification of the lens.<sup>158</sup> In addition, reports were not sent to consultants until at least ten of their patients had been followed up.

It should be noted that some consultants, such as orthopaedic specialists, were very happy with SF-36. Andrew Vallance-Owen and Brian Matthews, the project manager in charge of the PROMs project at BUPA, note that the specialties which were most positive about SF-36 were those that registered impressive health gains after three months.<sup>159</sup>

A related problem was the frustration of other hospital staff, who felt that the

153. Coulter A, Peto V and Jenkinson C, "Quality of life and patient satisfaction following treatment for menorrhagia", 2004

154. Connerney I et al, "Relation between depression and coronary artery bypass surgery and 12 month outcome: a prospective study", 2001

155. Vallance-Owen et al, 2001

156. Ibid

157. Ibid

158. Ibid

159. Vallance-Owen A and Matthews B, "Measuring clinical outcomes through patient surveys in independent hospitals. A presentation", 2001

anonymity afforded to individual consultants prevented them from identifying poor performers and protecting patients (and their hospital's reputation).<sup>160</sup> Here, a balance had to be struck between clinicians, who opposed greater openness of information, and managers, who wanted it. BUPA compromised by maintaining consultant confidentiality and developing a system to alert managers of any serious performance issues in their hospitals. The managers are then asked to investigate such issues further through other sources of audit data.

Another obstacle encountered by BUPA was that "questionnaire return rates were not sufficiently high to allow much reliance to be placed on the output".<sup>161</sup> Three years after BUPA's introduction of SF-36, records showed that 61 per cent of eligible patients received *and* completed the baseline form. There was considerable variation between hospitals, with more than a quarter showing a rate of over 80 per cent. The failure of hospital staff to entice their patients to complete the form (due in part to the frustrations noted above) has been cited as a major reason for the low completion rates. To improve the situation, a postal reminder was sent to patients who had not returned the questionnaire after three months. This raised the return rate to 75 per cent.<sup>162</sup>

Finally, some of the procedures being subjected to the measures did not contain sufficient cases to yield meaningful results.<sup>163</sup> While data was being collected on over 1,000 procedures, some of these only occurred once or twice in each three-month cycle. BUPA initially attempted to solve this problem by grouping together

procedures which could be expected to produce similar effects, but no such grouping ever met with enough confidence from some stakeholders. Consultants in particular worried that their results would be distorted by mixing procedures that could have very different rates of recovery. The solution eventually agreed upon was a narrowing of the measurement programme to 20 common procedures which covered the main surgical specialties. These included total hip replacement, CABG, adult tonsillectomy, hysterectomy and surgical removal of impacted teeth.

### Impact

There are specific examples of the data leading to problems of care being identified and addressed.<sup>164</sup> One BUPA hospital, alerted by the data to poor outcomes in one specialty, conducted further investigation and found the specific area of weakness to be post-operative pain relief for hysteroscopy. Targeted improvements were made. In another example, a consultant discussed his low ratings with his hospital's head of nursing and learned that the root of the problem was that he gave overly optimistic information to his patients about their recovery. He adjusted his approach accordingly.

"The programme", Vallance-Owen and his BUPA colleagues conclude, "has shown that clinical outcomes can be measured objectively and collected systematically using a comparatively simple process. It has also demonstrated that feedback must be presented in an understandable and user-friendly manner if it is to influence clinicians."<sup>165</sup>

160. Vallance-Owen et al., 2001

161. Ibid

162. Ibid

163. Ibid

164. Ibid

165. Ibid

---

# 7

## Outcome Measurement in Primary Care

### Summary

The NHS has pioneered quality measurement in primary care: it introduced a pay-for-performance scheme in 2004 and has invested £8 billion into primary care services over the past three years.

The introduction of the General Medical Services (GMS) contract was influenced by the ebb of trust from health-care professionals and the move to active monitoring of performance, which were discussed in Chapter one. The contract made headlines at the end of its first year because GP response exceeded that anticipated: practitioners claimed an average of 83 per cent of the available incentives for carrying out various treatments, whereas the Government had expected a figure of 75 per cent. As a result GPs' pay increased by 30 per cent, from an average of £80,000 to £106,000 a year, with some said to be earning as much as £250,000. The public have begun to question whether these generous rewards were matched by the gains in quality of healthcare. In January 2007, the Health Secretary Patricia Hewitt admitted that, with hindsight, the Government should have capped the amount that GPs could make from their new contracts.<sup>166</sup>

This chapter evaluates the criticisms that have been levelled against the GMS contract: that the evidence that it has improved quality is inconclusive; that the wrong things are being measured and rewarded; and that the design of the payment system encourages gaming and penalises practices serving deprived populations.

### Background

Before 2004 only a few measures of quality of care were routinely available. These were for services that attracted an additional fee, such as cervical smears, vaccination rates, child health surveillance, minor surgery and contraceptive surgery.<sup>167</sup> The GMS contract provides direct monetary incentives to the majority of general practitioners in the UK who are self-employed partners (or principals) and who share in the profit from their practices (including those opting for personal medical services agreements, the permanent local alternative to the national GMS contract). It awards payments, accounting for 25-30 per cent of income, for a wide range of services set out in QOF, which was developed from a number of different evidence-based schemes. The contract contains approximately 150 quality indicators across four broad areas: clinical, organisational, additional services and patient experience.

The relative payment for each indicator depends on a points system designed to reflect the likely workload involved. Each quality indicator is allocated a maximum payment and the monetary value of a point in turn depends on practice list size and demographics. For example, over half the maximum points were allocated to clinical performance (550 out of 1,050) and for an average practice with a patient population of 5,550 and three full-time partners the maximum payment for the clinical area alone was £66,000 per annum at 2005-06 rates.

166. GP pay rise 'was mistake', interview with Patricia Hewitt, 19th January 2007, see [http://news.bbc.co.uk/player/no/news/id\\_6280000/6280077.stm](http://news.bbc.co.uk/player/no/news/id_6280000/6280077.stm) (last accessed 03/05/07)

167. Ashworth M and Armstrong D, "The Relationship Between General Practice Characteristics and Quality of Care: a national survey of quality indicators used in the UK Quality and Outcomes Framework 2004-5. *BMC Family Practice*, vol 7, p 68, 2006

**Table 5: performance initiatives introduced before 2004<sup>168</sup>**

- National standards for the treatment of major chronic diseases, such as the National Service Frameworks for coronary heart disease (1999) and diabetes (2003)
- Contractual requirement for practitioners to undertake a clinical audit (initially a requirement in the 1990 contract)
- Financial incentives for cervical cytologic testing and immunisation (early 1990s)
- Widespread use of audit and feedback by the Primary Care Trusts
- Release of comparative data for quality of care to practitioners (common) and the public (rare) by the primary care trusts
- Annual appraisal of all primary care physicians (by the Primary Care Trusts and including discussion of some audit data)
- National system of inspection and monitoring of performance (by the Healthcare Commission)

The impact of the GMS contract on quality is unclear

A key question is whether the high levels of QOF interventions attained after the introduction of pay-for-performance reflect improvements that were already underway – an acceleration of existing trends – or a distinct response to a new incentive. Even if the latter were true, a subsidiary question remains: was it payment or measurement that encouraged the improvement?

The quality of primary care had already begun to improve before 2004 in

response to a wide range of initiatives, such as setting national standards for the treatment of coronary heart disease (introduced in 1999) and diabetes (introduced in 2003), and a national system of inspection.

The most up-to-date evidence from Martin Roland’s study of the quality of care for three chronic conditions – coronary heart disease, Type 2 diabetes and asthma<sup>169</sup> – suggests that the introduction of pay-for-performance was associated with a “modest acceleration in improvement” for diabetes and asthma. These results, set out in Figure 3, are based on care reported in the medical records and does not necessarily represent the care provided. A common criticism of pay-for-performance programmes is that their main effect is to promote better record-keeping rather than better care.<sup>170</sup>

The GMS contract looks at the “wrong” measures

*Too much process, not enough outcome*

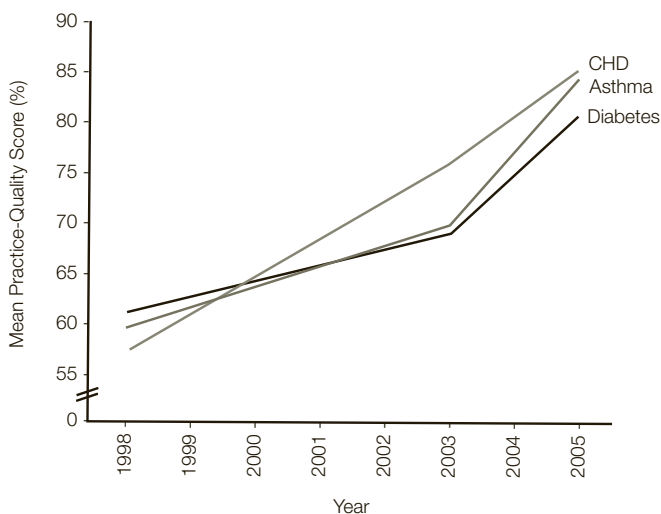
Despite its name the quality and outcomes framework is heavily biased towards measures of clinical process; there are only 20 individual outcomes measures in the QOF dataset of 150. A balanced mix of outcome and process indicators is essential.

168. Roland M et al, “Quality of Primary Care in England with the Introduction of Pay-For-Performance”, *The New England Journal of Medicine*, pp 181-190, 12th July 2007

169. Ibid

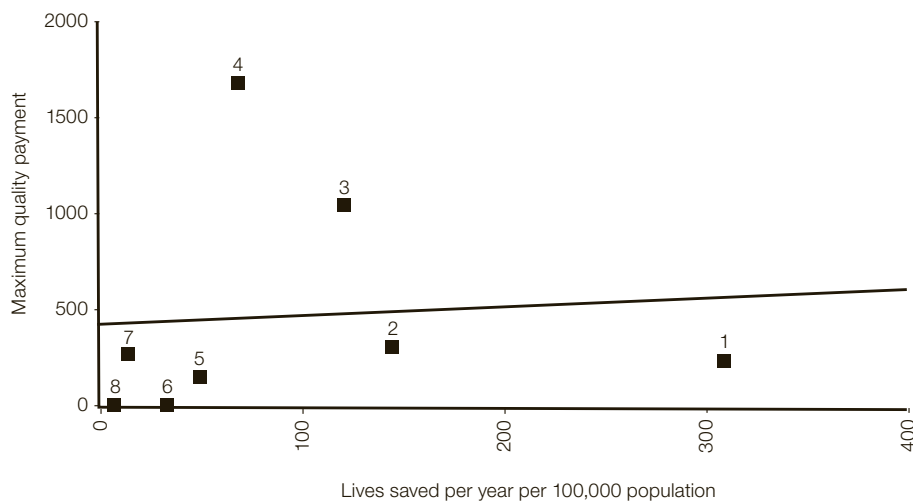
170. Rosenthal M and Frank R, “What is the Empirical Basis for Paying for Quality in Healthcare”, *Medical Care Research and Review* vol 63 (12), pp 135-157, 2006

**Figure 3: mean scores for clinical quality at the practice level for coronary heart disease, Type 2 diabetes and asthma, 1998-2005**



**Figure 4: potential quality payments against potential lives saved for the eight interventions**

Intervention (number relates to graph)	Maximum lives saved per unit of time	Maximum lives saved per 100,000 per year (% of total)	Maximum payment for typical practice
1 ACE in heart failure	76 per 90 days	308.0 (41%)	2400 (06%)
2 Influenza Immunization in over 65s	146 per year	146.0 (20%)	3600 (10%)
3 Stop smoking advice and nicotine replacement	120 per year	120.0 (16%)	10,440 (28%)
4 Screening and treatment of hypertension	286 per 4 years	71.0 (10%)	17,280 (46%)
5 Aspirin in Ischaemic heart disease	48 per year	48.0 (6%)	1320 (4%)
6 Warfarin in atrial fibrillation	33 per year	33.0 (4%)	0 (0%)
7 Statins in ischaemic heart disease	69 per 5 years	13.8 (2%)	2750 (7%)
8 Statins in primary prevention	14 per 5 years	2.8 (0%)	0 (0%)
<b>Total</b>		<b>742.6 (100%)</b>	<b>37,800 (100%)</b>



**Looking at the wrong interventions**

The principle underlying the points system is that it rewards GPs and their staff for the quantity of work done. The weighting of points was determined by two small groups of GPs who estimated the work required to achieve the different interventions. This approach – basing the rewards on perceptions of likely workload – has the advantage that it encourages GPs to give equal weight to all quality indicators, rather than prioritising the less burdensome ones. But likely workload may not reflect likely health gain.

Last year, Robert Fleetcroft and Richard Cookson assessed whether or not there is an association between improvements in health and the financial incentives of the GMS contract.<sup>171</sup> To do so, they examined eight preventive treatments covering 38 of the 81 clinical indicators in the quality outcomes framework (Figure 4). The maximum payment for each service was calculated and compared with the likely benefit in terms of lives saved per 100,000 population based on evidence from a widely endorsed study by McColl et al. Maximum

171. Fleetcroft R and Cookson R, "Do the Incentive Payments in the New NHS Contract Reflect Likely Population Health Gains?", *Journal of Health Services Research and Policy*, Vol 11, pp7-31, January 2006

payments for the eight interventions make up 57 per cent of the total maximum payment for all QOF clinical interventions.

Potential lives saved for the different interventions ranged from 2.8 to 308 per 100,000 population per year; potential quality payments ranged from zero to £17,280 per annum (at the 2005-06 rates). Fleetcroft and Cookson concluded: “There appears to be no relationship between pay and health gain across these eight interventions... There is a danger that clinical activity may be skewed towards high workload activities that are only marginally effective, to the detriment of more cost-effective activities.”<sup>172</sup>

Importantly, two of the interventions that have proven effectiveness – the use of warfarin in atrial fibrillation and statins in primary prevention of heart disease – received no quality incentive payment at all. This is in contrast with other areas, such as personal learning plans, that receive incentives but whose benefits are unsupported by robust evidence.

#### *Distorting priorities*

Rodney Haywood has also argued that guidelines for treatment and outcomes, when used for performance measures, may encourage increased intervention and expenditure with little evidence to support their effectiveness. Such measures often become bargaining chips for powerful special interest groups. “Influential parties often have strong incentives to advocate that these measures be aligned with idealised goals... even the most pure-hearted persons and groups with vested interest in issues related to diabetes... have a natural and justifiable tendency to want more attention and resources for their cause.”<sup>173</sup>

At the same time other disease areas – such as mental health, learning disabilities, palliative care of the elderly, osteoarthritis, osteoporosis and multiple sclerosis, where successful outcomes are harder to measure, or whose advocacy groups have less power

to press for “idealised” measures of success – are neglected. What is not incentivised is marginalised.

#### *Local versus national priorities*

A further criticism of the GMS contract is that it lacks the flexibility to address local health issues, especially pockets of deprivation.

Ethnic minorities often suffer more than the majority population from conditions such as diabetes. Christopher Millett et al analysed outcome measures that indicate high-quality diabetes care, such as patient levels of glycated haemoglobin (HbA1c measures the extent to which diabetes is under control), in the London Borough of Wandsworth. They concluded that the introduction of pay-for-performance had not addressed disparities in the management and control of diabetes in ethnic groups.<sup>174</sup> Although this particular study somewhat overstates its case, it is nevertheless true that the easiest way of scoring high may be to target the easiest to reach groups, thus increasing inequity. For example, if 90 per cent is the cut-off point for maximum payment, there is no incentive to seek out the 10 per cent hardest to reach – and most expensive – patients.

Of course, reducing these type of health inequalities would not be a priority in areas with smaller minority ethnic populations and, for this reason, the possibility of some local discretion in setting a measure of “success” is attractive. Multi-ethnic Wandsworth might want to include measures of equality of access and outcome, other local health services could address age or socioeconomic inequality, and so forth.

#### Unintended consequences

##### *A high level of gaming?*

As Tim Doran has written: “Evidence-based quality indicators should not be applied unthinkingly, since patients have

172. Ibid

173. Haywood R, “Performance Measurement in Search of a Path”, *The New England Journal of Medicine*, p.951, 1st March 2007

174. Millett C et al, “Ethnic Disparities in Diabetes Management and Pay-for-Performance in the UK: the Wandsworth prospective diabetes study”, *PLoS Medicine*, June 2007

co-existing conditions that affect their optimal care. It is inappropriate, for example, to strive to control the cholesterol level of someone terminally ill with cancer.<sup>175</sup> Clearly, intrusive tests on dying patients are both inefficient and unethical. The GMS contract recognised this and allowed GPs to exclude patients from specific treatments (“exception reporting”). However, this provided an opportunity for practitioners to increase their income by excluding patients inappropriately.

A small number of practices appear to have achieved their high scores in the first year of the contract by excluding large numbers of patients. More than 1 per cent excluded over 15 per cent of their patients.<sup>176</sup> It is possible that those practices which were better at identifying and treating patients were also better at identifying patients for whom the targets were inappropriate, and the low level of exception reporting suggests that gaming on a large scale was rare.

The rate of exception reporting varied considerably by disease group; this may reflect variation in the ease of meeting individual targets or the amount of financial reward available for each. For example, there were low levels of exception reporting for hypothyroidism (worth about £456 in the first year of the contract) and high levels for mental health problems (worth about £1,748), coronary heart disease and chronic obstructive pulmonary disease.<sup>177</sup>

Has the new contract created widespread gaming throughout primary care in the NHS? Probably not. But the very high levels of exclusions in a minority of practices suggest that the capacity to game exists and is being exploited. There is also the danger of double counting – by including care provided in hospitals or from district nurses employed by primary care trusts.

### *The GMS contract and deprivation*

The option to exclude patients is not the only reason why the contract does not ade-

quately reward practices that seek out the neediest patients. During the final negotiations on the contract in 2004, it was decided to take into account the number of patients in a practice with a particular disease in order the better to match payment to workload. But rather than use true prevalence of disease, an approximation, the adjusted disease prevalence factor (ADPF), was formulated. Although the calculations under ADPF are complex, the main aims, wrote Bruce Guthrie were to “reduce variation [in payment] and relatively protect the losers, while at the same time providing fair rewards to those who have the highest prevalence”.<sup>178</sup>

Evidence suggests that the ADPF does not reduce variation in total practice income: there is a forty-four-fold variation in payment for practices treating the same number of patients to the same level of quality.<sup>179</sup> The aim of fair pay for workload is therefore not met. Moreover, the ADPF institutionalises what has been termed the “inverse care law”: the tendency for more funding to go to areas that have less need of it than others. This payment does not achieve its main goals and illustrates the problems that occur when complex adjustments are introduced too rapidly and without adequate pilot trials.

Variability in the quality of care offered by different practices has long been a cause for concern and overall QOF scores have already been found to be lower in areas of social deprivation. Could this be explained by the difficulty of providing good quality care to the neediest populations? Could a link between social deprivation and poor primary care quality scores arise because they have larger, more unmanageable lists, or because they have a higher turnover of patients making it difficult to accumulate sufficient clinical success? Mark Ashworth and David Armstrong investigated and found that “a high proportion of patients aged 75 years or over or a high proportion of the local population being born in a

175. Doran T et al, “Pay-for-Performance Programs in Family Practices in the United Kingdom”, *The New England Journal of Medicine*, pp 175-384, 27th July 2006

176. *Ibid*

177. *Ibid*

178. Guthrie B et al, “Workload and Reward in the Quality and Outcomes Framework of the 2004 General Practice Contract”, *British Journal of General Practice*, November 2006

179. *Ibid*



180. Ashworth M and Armstrong D, "The Relationship Between General Practice Characteristics and Quality of Care: a national survey of quality indicators used in the UK Quality and Outcomes Framework 2004-5, *BMC Family Practice*, vol 7, p 68, 2006

developing country did not have an independent effect on the QOF score." They also found no evidence of lower QOF scores with very high list sizes. Similarly, "high turnover of the registered practice population appeared not to make it more

difficult to achieve higher QOF scores".<sup>180</sup> In other words, the researchers did not find that deprivation of itself caused poor quality primary care.

---

# 8

## Recommendations

Summary: what can and should we do?

We began this report by noting that NHS professionals have traditionally been self-regulated and have had professional autonomy. This model was rejected in the 1990s due to poor health outcomes, long waiting times, and a series of scandals involving incompetent or malign clinicians. The Government's response to this was the development of national standards of performance management. These included clinical governance and monitoring by the CHI, and a massive increase in public spending on the NHS. Some improvements in care have been achieved; for example the NHS now has shorter waiting times for virtually all points of access to healthcare. In other areas, however, significant problems persist; England still suffers from relatively poor outcomes in many areas of care, and the regulatory framework and systems of monitoring, accountability and regulation are confusing and underdeveloped, and hence often ineffective.

In England, between 1997 and 2002, the Government sought to ensure that patients would have access to good local services. A system of star ratings held chief executives of NHS organisations to account for the delivery of Government priorities, while CHI undertook a comprehensive rolling programme of inspections. Since 2002, the Government has sought to ensure better access in a pluralistic system, including NHS trusts, foundation trusts and private providers. The Government has tried to facilitate patient choice between competing providers on the basis

of quality. However, the arrangements for, and underpinning of, this pluralistic system have caused problems:

- The Healthcare Commission is tasked (and funded) to undertake inspection of healthcare services on a targeted and proportionate basis. Therefore it is not sufficiently sized to undertake a comprehensive rolling programme of inspections (including visits) of the complete variety of providers that are publicly financed (which due to the encouragement of new entrants into the NHS 'market' is massively increased).
- Given the limitations on resources, the majority of Healthcare Commission activity in relation to the NHS has been focused on its statutory obligation to produce an annual healthcheck that assesses the quality of NHS providers. However, this is limited in two ways. First, the parameters of the healthcheck are largely determined by Government priority (targets new and existing, and minimum standards for NHS care) and second, the system of assessment must, by statute, apply to the whole organisation. This creates a weakness; important dimensions of quality that lie outside Government targets, or cannot be related directly to minimum standards, cannot be reflected adequately in the healthcheck.
- A second limitation that current statutes place on the Healthcare Commission is a different legal set of requirements, and hence different regu-

latory and assessment regimes, for the NHS and the independent sector. In a system which is supposed to be pluralistic and decentralised, it makes no sense to judge the same service provided by the NHS and independent sectors in different ways.

“ Information is often seen as facilitating patient choice, but it also serves activation, the patients’ willingness to be involved and assertive in the decision-making about their own care ”

- It is unclear how two different policies to improve quality relate: inspections by the Healthcare Commission and patient choice. To drive up quality through patient choice, patients ought to have specific and timely information on services – not just information at the aggregated level of the hospital. While the Healthcare Commission does undertake service specific assessments, these have been of services such as mental health or chronic conditions which are less susceptible to improvement through competition (and where, therefore, regulatory intervention is a more pressing need). Therefore there is a paradox that the areas on which a regulator should properly concentrate its resources are precisely those areas where publication of performance is least likely to stimulate patient choice.
- We lack evidence to support the idea that patient choice is a driver of quality of care. Indeed, the evidence from the US is that patients do not always use this information when it is available. Moreover, our own survey shows that patients want information at the aggregated level of a hospital (which is a weak indicator of quality), primarily for the purposes of reassurance rather

than choice. Research by the Picker Institute<sup>181</sup> also shows that patients have little interest in being offered a choice about which hospital they would like to be admitted to. It may be that over time patient choice could become a key spur to quality improvement in healthcare. However, it has not yet reached that stage.

A future information ‘landscape’

In this study we have made a case for collecting more information. We have focused on practical examples of good use of information and explored international examples of best practice. In order to improve the quality of care we suggest that information should be used in three ways:

**1. Information for accountability**

Accountability measures are designed to assure taxpayers that their local health services (including NHS providers, Independent Treatment Centres and private providers doing publicly funded work) do two things:

- provide at least a minimum acceptable level of care, including safety, access, clinical competence, and compassion;
- spend public money with due care and consideration and in order to achieve the goals which the Government decide are appropriate.

**2. Information for patient choice and activation**

Information is often seen as facilitating patient choice, but it also serves activation, the patients’ willingness to be involved and assertive in the decision-making about their own care.

**3. Information for providers (for quality improvement)**

This information is used to activate providers’ intrinsic motivations of professionalism and altruism. It does not neces-

181. Available at <http://www.picker-europe.org/> (Last accessed 03.05.07)

sarily need to be published but does require a mechanism to share the information with providers and hospitals, and to encourage and monitor improvement.

Information for accountability

#### *Organisational level*

Accountability does not mean that the hospital guarantees that nothing will ever go wrong anywhere within its walls, but that it will do all it reasonably can to ensure a basic level of care. Therefore, the measure should be at the level of organisation, such as hospital trust, rather than sub-organisational. This is the only one of the three types of measures that should be at this high level of aggregation.

#### *Nature of data*

The data will consist of a mixture of judgements made by competent professionals following a rigorous and consistent process of assessment. An example would be the Healthcare Commission's judgement of core standards, which utilises pre-existing data, and inspection, by qualified professionals.

#### *Who would gather, analyse and publish the data?*

The Healthcare Commission is ideally placed to undertake this role. It already collects, analyses and publishes much of the data required for the purpose of accountability through its annual health check. The current responsibility of the Healthcare Commission to create an individual rating for each institution obscures some of the value of the data, but better organisation and presentation of the information should be relatively straightforward. Another advantage of using the Healthcare Commission is that proper accountability requires independence. The NHS performance management process cannot substitute for outside scrutiny. Equally, private companies which also pro-

vide consultancy services for NHS trusts cannot play this role, because of a clear conflict of interest.

#### *Presentation and dissemination*

The information should be disseminated through Healthcare Commission publications and mass and local media. The idea of requiring hospitals to display their current assessment within their premises should also be considered. Presentation should be targeted to the specific questions identified below, and, to make the results easily understood, should be presented in a way that shows "better than expected/high quality", "as expected", "below expectations/ some concerns".

The accountability matrix in appendix two illustrates some measures that could be used.

Information for patient choice and activation

#### *Purpose*

Information is often seen as facilitating patient choice, but it also serves activation, the patients' willingness to be involved and assertive in the decision-making in their own care. Moreover, choice can only really apply to predictable elective care, and especially surgery. It is not generally relevant to emergency care or to management of long-term conditions, where it will tend to conflict with continuity of care.

Whether for choice or activation, releasing small amounts of information about the quality of specific services to patients with specific conditions rather than consumers generally will provide invaluable support.

#### *Organisational Level*

Information aimed at improvement needs to be at the level of the service: practice, not PCT; department, not hospital.

#### *Nature of data*

The data will generally be individual measures of performance which have been suffi-

ciently tested for robustness and usefulness. This type of data needs greater specificity than information for accountability, so judgements are less appropriate than measures.

Earlier in this report, we proposed a classification of measures which covered structures, processes from the point of view of doctors and patients, and outcomes both from the point of view of doctors (measures of mortality and morbidity) and patients (such as the self-reported return of physical function). All of these, except structural measures (vital for organisations to diagnose the causes of their level of performance but of less immediate value for patients), are suited to providing information for patient choice and activation.

The data required to support these measures are generally available (and are likely to become more so in future) with the exception of PROMs, which are a major omission from current NHS datasets. BUPA have shown that it can be done through their extensive use of the SF36. The cost of this data collection is estimated at between approximately £3 and £15 per patient, which would equate to no more than 2 per cent of the 2007/08 Payment by Results tariff for a cataract operation and no more than 0.25 per cent of the same tariff for a knee replacement. The gains to be made from using such measures, and the low cost of doing so, make the case for their introduction into the NHS very persuasive.

In terms of the presentation and design of the data, US-based research suggests that:

- information should be focused to allow the patient to make judgments about specific elements of care.
- comparison should be made with norms and acceptable levels of performance, rather than trying to create spurious league tables of providers.

### *Who would gather, analyse and publish the data?*

There are multiple roles to be played in delivering these types of data, and different players from the voluntary, private and professional sectors are able to play them. As with accountability measures, trust is essential for the measures to be valued but, unlike them, this need not require the impersonal formality of a regulator. For one thing, the gaps in what measures are currently available are simply too great for one regulatory body to fill quickly. We propose that as many organisations as meet required standards of technical expertise to produce measures to a high standard of quality should be encouraged to use national data sources to do so.

The issues to be considered when developing these measures are as follows:

- Determining *what* to measure. Our classification of information types identifies two points of view, that of the patient and that of the clinician. There are two groups of bodies particularly well placed to determine what is measured: disease-specific patient groups (e.g. Diabetes UK) and clinical specialty groups (e.g. Royal College of Cardiac Surgeons).
- Determining *how* to measure and analyse the data. This often requires considerable analytical skill and technical expertise. This is located in a range of public and private sector organisations, such as the NHS information centre, the Healthcare Commission, several university departments, the King's Fund, CHKS, and Dr Foster Intelligence to name just a few.
- Data collection and management. This divides between quasi-administrative data such as HES, which is the responsibility of the NHS information centre; quasi-regulatory data, which is collected and managed by the

Healthcare Commission; and clinical audit type information, which is collected and managed by Royal Colleges.

- Publication and dissemination. Patient groups are ideally placed for this role as they know who the likely beneficiaries of such information are through their networks of membership, they are a trusted source of information, their dissemination channels are already established and they have a specific skill and credibility in communicating to their target audiences.

There are many potential models of how this information could be provided. One would be for the Government to support, through “seed money” and guaranteed access to publicly collected data, a set number of projects each year to be led by patient groups to supply information about performance for their specific audience. These groups would be allowed to commission the analysis and data management from a range of approved provider organisations, such as those listed above. Clearly, some entry requirements for providers of analysis would be necessary in order to ensure information of sufficient quality, but this decentralised approach is likely to ensure quicker and less bureaucratic production, easier targeting of information at the right audience, information that comes via a trusted “brand” for the audience, and quicker filling of the gaps in knowledge about healthcare.

The patient choice and activation matrix

Appendix three illustrates an example information matrix for diabetes, drawing, wherever possible, from existing data sources. These measures should be summarised into assessments into the four areas at the “above/as/below expected levels”.

Information for improvement

#### *Organisational Level*

Information aimed at improvement needs to be collected and presented at the level at which changes need to be made (if necessary clinical team or practice)

#### *Nature of data*

There are two potential models of the type of data which could be used in this context: the *surveillance model* or the *comprehensive model*.

#### *Surveillance model*

This model uses a range of outcome measures derived from routinely collected data sets which can be used to identify potential areas of weakness (for example mortality rates derived from HES and to stimulate internal analysis and improvement activities. These may either be a wide range of related measures where a consistent pattern of relatively poor performance would raise concerns, or it may be sentinel measures of serious failure (e.g. death following day case surgery).

Typically these will be time series data which spot quite small but significant changes in performance over time. These include statistical techniques such as CUSUM and control charts. At their best they can run in almost real time, allowing quick response to problems before they become crises. This is clearly a cheaper option than the comprehensive model, but has the risk of missing failing services, or of not picking up weak processes or structural failure before bad outcomes result.

#### *Comprehensive model*

The alternative model is to create more complete networks covering all aspects of the structure, process and outcome measures from both clinical and patient perspectives, and use these effectively for internal performance management and diagnosis of issues. At their best, comprehensive models would combine the best

elements of good clinical audits, patient experience surveys and PROMs. So this is a gold standard, but expensive, option.

A rough estimate of the cost of implementing this model nationally, based on HC Improvement Reviews, which are somewhat similar, would be a minimum of £4 million. The most onerous costs an absolute minimum of £4 million to implement *nationally*, of which the most significant costs are those of local data collection (this figure excludes costs associated with improvement activities). More fundamentally, the staff needed to undertake this sort of work are not available in the NHS, unless there is a fundamental re-structuring of the job role of NHS informatics staff away from unnecessary data collection for central reporting and business management purposes.

*Who would gather, analyse and publish the data?*

There are a range of different players here, all of whom have slightly different roles, and all of whom, we argue, are needed:

*Internal performance management*

The failures in quality which were discovered in the 1990s would have been taken more seriously if they had been concerned with financial management. Performance management in the NHS, despite the commendable intentions behind the introduction of clinical governance and core and developmental standards, still concentrates on money and access rather than clinical quality.

Judging quality is not straightforward, and performance management against such judgements still more complex. But NHS management is now reasonably well structured to take on this role, particularly if the comprehensive model is to be used. Strategic Health Authorities (SHAs) are large enough bodies to collect, analyse, and, critically, compare data to allow identification of outliers of good and bad performance. They have estab-

lished management relationships which allow both accountability of performance and the capacity to share good practice between different providers (in the manner of the Veterans Association in the US). The local element is also important. The priorities for improvement are likely to be different in different parts of the country, and since SHAs operate below national level, they have the capacity to set quality improvement priorities which more closely reflect local conditions.

There is an emerging consensus that a multi-provider NHS (at least in some sectors) will emerge over the next five to ten years. In this context the commissioning role of PCTs is essential. Given that the payment-by-result tariff limits choice on price, commissioning decisions should in theory be made on quality. At the moment, the information to support PCTs in doing this is not available. Measurement of quality by SHAs would create a source for this information. It would also further strengthen the incentives to improve inside the SHA performance management process.

*The external regulator*

The Healthcare Commission has two distinct programmes which relate respectively to the surveillance and comprehensive models. The first is a screening and surveillance function, and the second is “improvement reviews”, which look in detail at individual services.

The commission is able to do this as in addition to its analytic functions, it has a large field force of inspectors (known as assessment managers) who can act as translators of this information to the services that are being inspected.

There are, however, limits to what a regulator can legitimately do in this field. These relate to its size, role and function. At its current level of funding (equivalent to 0.07 per cent of the total health expenditure in the UK) it is simply too small to

undertake comprehensive assessments of all areas which need to be covered. If we expect competition to improve quality, the regulator should limit its improvement-focused activities (as opposed to those ensuring a minimum standard or providing information to facilitate improvement by others) to those areas where “market”-type solutions are less likely to work.

Finally, there are considerable philosophical problems in involving a regulator (or indeed a quasi-regulatory body) in improvement activities. The result could be a confused relationship with the bodies being regulated. A regulating body could find itself criticising its own activities, which would obviously be difficult.

All this points to the surveillance model as a more attractive general approach for a regulator to use: it is cheaper, more locally responsive, and more focused, quickly identifying and correcting poor performance. To be used properly however, it needs to link with the general performance management framework of the NHS. For example, the regulator should use its field force to address a problem which exists at only one institution, but where the problem is endemic across the area, broader performance management processes for organisations across the area are more appropriate.

#### *The external consultant*

There are two types who can be involved. There are the private sector companies who have a strong analytic, information management and presentation focus. Obvious examples include Dr Foster and CHKS. For example, Dr Foster have developed and used a surveillance type of system, and their “intelligent board” reports specify data sets which bear some resemblance to the comprehensive model described above. However, there is almost certainly room in the market for analytic specialists working on specific local problems, and indeed very often such small,

targeted, quick and cheap responses would be far more appropriate to a local situation than more general approaches from larger companies.

The other type of consultant is expert in performance improvement methodologies such as “Lean”. Again, interventions by such consultants can be entirely appropriate in certain situations. One would, however, expect them to work with analytic experts and to become involved in response to the data.

We believe that the involvement of both groups should be encouraged, as the whole range of potential issues in health-care is too wide to rely on performance management and regulation alone. In addition, such organisations can respond more quickly to local problems. They do not have other statutory responsibilities, nor a broad societal/utilitarian remit that must apply to public bodies and which mitigates against analysing unusual issues. We believe that the impetus for using their services should come from individual providers seeking to solve their own issues, rather than being imposed by central government or regional management.

#### Recommendations

- 1 The Government should explicitly recognise three different uses for information: accountability to the public, choice and assertiveness for patients and improvement for providers.

#### *Accountability*

- 2 The Government should explicitly reconfirm the role of the Healthcare Commission to publish information about minimum acceptable standards of care and value for money for general public consumption.
- 3 To assist this, the Government should remove the requirement that the Healthcare Commission must give an



- overall rating for every trust every year, and replace this with a constantly updated “accountability balance sheet”, as described in appendix one.
- 4 Hospitals should be required to display their current performance as it relates to safety, accessibility, competence and compassion.
  - 5 The importance of public health is now recognised across the political spectrum, and the new national targets focus on this is to be welcomed. However, holding individual healthcare organisations to account for achieving them makes little sense, and has led to a situation where the indicators used to judge performance have only a tangential relationship with the ultimate goal. We suggest that rather than holding individual healthcare providers to account, performance at a SHA level would allow more meaningful measures to be used.
  - 6 QOF data is currently released on the internet by the Information Centre for Health and Social Care, but its presentation as an abstract point-scoring system makes it difficult to understand and evaluate. There should be more transparent release of data.

Practices should provide aggregated QOF data for key performance categories (cardiovascular, respiratory, mental health, patient satisfaction and management). The measures in these groups should be decided by independent representatives after public reports of outcomes evidence. Data should be updated every month.

#### *Targeted patient information*

- 7 Information should be collected and published at the appropriate specialty or disease group level.
- 8 The Department of Health should expand funding for National Clinical Audits, but mandate that the information should be made available (at a suitably aggregated level) for publication and use by interested parties. As a minimum, three further disease/procedure specific websites (of the manner of the heart surgery website) should be created within five years.<sup>182</sup>
- 9 The Department of Health should run a pilot scheme for providing seed money to patient groups (e.g. Diabetes UK) to develop and publish a “patient-focused” report for individual services using suitably qualified analysts and immediately available data with a view to introducing this as an ongoing commitment to openness.
- 10 The long-term aim should be to have good enough high quality data collected, expert analysts available, and procedures established so that in five years time any patient interest group can use nationally collected data to produce information for their members which is accurate, helpful, authoritative, scientifically meaningful and easily understandable.

#### *Information for performance management and quality improvement*

- 11 SHAs have a critical role in using NHS performance management processes to increase quality of care. We advocate allowing SHAs to set local goals and incentives to achieve this. This would require Government to acknowledge that variation across areas of the UK in the priorities (and quality) of different services is a legitimate response to the variation in population health needs and priorities.
- 12 SHAs should be required to develop and implement “comprehensive model”

182. <http://heartsurgery.healthcarecommission.org.uk/> (last accessed 03.05.07). We believe that responsibility for the publication of such information could lie with any one of a number of organisations - patient groups, royal colleges, private sector information providers or the health services regulator.

reviews of the quality of one of their services each year, which should generate organisation-specific improvement goals for the next year. These should focus on areas which are of specific local concern.

- 13 The Healthcare Commission should continue to develop its use of “surveillance models” in order to minimise the burden of collecting data and target its attentions on poorly performing services. It should, in particular, focus on areas where patient choice mechanisms and plurality of provision cannot easily be used.
- 14 The high levels achieved during the first year of the GMS contract suggest that its targets were too easy and there should be an increase in the thresholds for point scoring. The Government attempted to address this issue by raising the payment thresholds, modifying or dropping 30 indicators and adding an additional 18 for 2006-07.<sup>183</sup> We welcome the increases in the payment thresholds and believe that further rises should follow in 2007-08. A system in which 83 per cent of available points are claimed is scarcely very discriminating; progressively tougher thresholds should allow greater discrimination between high and low performers, as well as encouraging improvement.

*Information collection and analysis – the supporting infrastructure*

- 15 In order to fill the gap in knowledge about the outcomes of most medical interventions, the Government should invest in SF-36 for at least two surgical specialties across the whole NHS in the next year, as an interim measure. Longer term, involvement with the still more robust PROMIS project should be sought. More outcome rather than process
- measures should also be introduced into the GMS and PMS contracts. This would help to quell the criticism that these contracts have given GPs a large pay rise for better recording of activities that they were already undertaking. Options for doing this include creating patient reported outcomes measures (PROMS) for those undergoing management of chronic conditions and measuring unplanned, preventable use of hospital services.
- 16 The adjusted disease prevalence factor has not succeeded in focusing resources on areas of high morbidity. The ADPF should be replaced by the TDPF. For practices with the same number of patients on the register, payment additionally increases with the size of the disease list.
- 17 NHS Information Centre data should be made easily available (including easy to use analytic tools) to a wider range of users (which includes NHS information departments) in order to encourage a market in analytic services for the NHS - this should include NHS information departments and private sector analytic consultants.
- 18 Government should ensure that service level agreements between commissioners and providers include timely discharge data prepared in accordance with national standards set by the Information Centre.
- 19 In order to make this successful, the NHS Information Centre should design criteria for minimum standards to be maintained by users of their information – these should concentrate on knowledge of the data, analytical skill and organisational capacity (for

183. Negotiators agreed that all lower thresholds for existing indicators should be raised to 40%. The upper threshold will remain at 90% for the majority of indicators. For those with an upper threshold of less than 90%, it will be raised in line with UK average achievement in 2005.

New clinical areas agreed include: heart failure, dementia, depression, chronic kidney disease, atrial fibrillation, obesity and learning disabilities. The following clinical indicators sets were amended: mental health, asthma, stroke and transient ischaemic attack, diabetes, chronic obstructive pulmonary disease, epilepsy, cancer and smoking

example to respect data protection legislation).

20 In the long-term, the national NHS IT system (Connecting for Health) “spine” provides excellent opportunities for enhanced data collection and use. To start planning for the more effective use of this resource, we endorse the recent Conservative party white paper’s call for the establishment of a “reference group of academics, economists and the professions” and recommend that once the practicalities of this group’s working are finalised, item one on their agenda should be to consider how best

to exploit the data that the NHS IT system will contain for better measurement and improvement of quality of care.

21 We recommend caution in using any further financial incentive systems to increase quality. Any future developments should be piloted and properly studied to understand potential perverse effects and potential for manipulation before being run nationally. The pilot should include a control group of organisations receiving information about the incentivised performance measures but no financial incentive.

---

## Appendix one

### Explaining why US comparisons are valid

#### *Summary*

US healthcare is structured very differently from the NHS. It is characterised by a multiplicity of providers and purchasers; greater emphasis on individual choice and less on universality and equity; greater fragmentation of services; and substantially greater cost. Whatever its shortcomings as a system, publication of outcome data is more entrenched in the US.

American case studies are extremely valuable in any attempt to make the case for outcomes data in this country. The US is the world leader in publishing information about clinical outcomes, a status that reflects its culture of individualism, choice and mistrust of paternalist authority – characteristic of a young country with no experience of aristocracy or feudalism.

Another explanation is that the rapidly increasing cost of healthcare has led to greater efforts to ensure value for money, of which measures of quality are an important part. The US has by some margin the most expensive healthcare system in the developed world, accounting for some 15 per cent of GDP.

Part of the problem is a payment system that, in general, pays doctors by the procedure. This, combined with an increasing perception of the likelihood of litigation resulting from any error or adverse outcome, encourages clinical padding – doctors performing more tests and interventions than are strictly necessary. Rewarding doctors on the basis of actual improvements in health as a result of their care, rather than paying for all the interventions they make, can restrain the growth of these needless costs. Employers, major purchasers of healthcare for their workers, are also keen to maximise value. Rather than

focus exclusively on premium costs, they want improved health outcomes for their workers in order to reduce health-related absence.

This is not to suggest that it is only supply side concerns that have driven the growing interest in outcome measures. Facilitating consumer choice by allowing patients to be better-informed about healthcare providers is still the ultimate justification for measuring quality. However, as we will see later in this section, the evidence that patients pay attention to such data is limited. The more realistic expectation is that providers themselves will be moved to improve their performance, almost regardless of whether patients drive the process through their purchasing decisions.

The complex and fragmented nature of the US health system can be confusing for British observers used to the simpler monopoly of the NHS, and make it difficult to transfer practices between the two countries. There is a plurality of providers and purchasers, allowing for theoretically wider choice and more immediate access, at least to specialists. Primary care, very much the poor relation in the US, does not play the same gate-keeping role that it does in the NHS. Indeed, attempts to create a more integrated, coherent and cheaper system, such as the introduction of Health Maintenance Organisations, have incurred resistance precisely because they are seen as rationing care and limiting choice. At the same time, the wider range of providers leads to greater fragmentation of services which increases transaction costs. This is often met with micro-management of clinical decision making to reduce costs, whilst

clinicians often need pre-approval from insurance companies for certain procedures and referrals.

The US model is easily caricatured as a private health insurance system, with insurance typically linked to employment, but this picture is somewhat misleading. The true nature of the American system is more complex, with massive state intervention in the form of Medicare and Medicaid. Indeed, public spending on healthcare as a share of GDP is as high in the US as it is in the UK. At the same time employment-linked health benefits are starting to crumble, with fewer employees receiving coverage as part of the package, and many of those that do suffering reduced benefits that often require high out-of-pocket expenditure.

Finally, there are estimated to be 45 million Americans without health coverage. Some of these either choose to make out-of-pocket payments for their healthcare or are only uninsured for a brief period, but many can only receive treatment by turning up at emergency wards, which are legally prevented from turning them away but may still charge them for part of their treatment. In all, despite its progress in publishing high quality information, the American health system may be the least attractive in the developed world, combining extravagantly high costs with serious problems of coverage. This report should

not be read as advocating a general shift towards the American way of organising healthcare.

As Marshall et al observe, “Public reporting in the United States is now much like healthcare delivery in that country. It is diverse, is primarily market-based, and lacks an overarching organizational structure or strategic plan. Public reporting systems vary in what they measure, how they measure it, and how (and to whom) it is reported.”<sup>184</sup> Organisations involved in reporting include the National Committee for Quality Assurance (NCQA), the Centre for Medicare and Medicaid Services, the Leapfrog Group, the Consumer Assessment of Health Plans, and Healthgrades.

### *Conclusion*

This brief overview has demonstrated that, for reasons of culture and cost-limitation, the US has led the way in measuring outcomes in healthcare. It should be said that there is no reason why the lead enjoyed by the US over the UK in this field should be permanent. A drawback of America’s fragmented system is the absence of a single, rationalist system for collecting information. The data is, as a result, patchy and often based on samples of patients. It is also difficult to refine as different organisations measure in different ways. The UK, with perhaps the most unified health structure in the world, does not face these difficulties.

184. Marshall M, Shekelle P, Davies H, Smith P, Public Reporting on Quality in the United States and the United Kingdom, *Health Affairs*, 2003

## Appendix two

### Information for accountability

#### (a) Quality of care matrix

Area	Data	Sources	Data type	Data Available/ Analysis undertaken	Additional cost
Safety	Judgement of meeting safety domain of core standards	Pre-existing large data sets (e.g. Hospital Episode Statistics (HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
	Measures of infection control – e.g. MRSA and clostridium difficile exposures	Special collections and Health Protection Agency data	Outcome measures	Collected in part, analysis somewhat limited	High
	Investigation (HC, Royal Colleges, GMC)	Special collection	An "over-riding" judgement	organisational failure should over-ride other findings on safety	Low
Accessibility	Judgement of meeting accessible and responsive domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
	Waiting times	Currently collected data	1. Measurements based on breaches of targets 2. Progress over time in reducing waits 3. Consideration of the whole distribution of waits (e.g. percentage of patients seen in 3 months, in 1 month etc)	The data exist to produce a wider range of measures which give a fuller picture of performance on waits and indeed are publicly available but are poorly presented and not publicised.	Low
	Avoidance of waiting time perverse effects	HES, "Suspension" rates, used to target inspections where there are concerns that either reported figures do not reflect reality, or where there may be perverse consequences elsewhere in the delivery of care.	Judgement based on inspections targeted by analysis of routinely collected data	It covers gaming, poor data quality, and unintended consequences.	Moderate
Competence	Judgement of meeting clinical (and cost) effective domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
	Patient Survey assessment of e.g. confidence and trust in doctors treating patients	Patient Survey	Outcome from patient perspective	There needs to be additional work either to identify the key few sentinel measures that represent a broader view of performance OR to find a method of combining and summarising these data.	Moderate
Compassion	Judgement of meeting patient focus domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low

Appropriate spending  
 This can be subdivided into three areas: measuring of general good governance, measuring of best practice and measuring of societal value. We suggest that this could be achieved relatively cheaply and straightforwardly by building on the work that the

Healthcare Commission is already doing, but this means that the Healthcare Commission should be given greater freedom to judge the manner in which it presents information, and should not be compelled to create an annual rating for each healthcare organisation.

**(b) Appropriate spending matrix**

Area	Data	Sources	Data type	Data Available/ Analysis undertaken	Additional cost
Good governance	Audit Commission/ Monitor assessment of financial management etc	Audit Commission/ Monitor	Judgement based on formal audit process	Yes	Low
	Judgement of meeting governance domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
	Evidence of poor use of data	Analysis of routine data sets for poor data quality and anomalies that suggest gaming of data collection and recording to target inspection	Judgement based upon structure, process and outcome data – should have an overriding effect if dishonesty can be demonstrated	Requires development of analysis and investigation techniques	Moderate
Use of best practice	Acute hospital portfolio	Specially collected data reflecting use of most cost effective approaches	Largely Structure and process data	Only available for hospital sector. Each review run rather as an individual project, so any new areas have specific development costs, pre-existing measures have collection costs	Moderate
	Judgement of meeting (clinical and) cost effective domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
Societal value	Judgement of meeting public health domain of core standards	Pre-existing large data sets (e.g. HES) to target inspection	Judgements based upon structure, process and outcome data	Yes	Low
	Performance against new national standards	Special collection	Largely structural data	We believe that public health measures should be collected at a locality level (e.g. an SHA area) where meaningful action can be pursued.	High
	Performance against existing national targets (e.g. waiting times in A&E departments)	Pre-existing collection – minimal analysis	Primarily process data	These are now well established, but inside this model the relative importance of these measures is reduced substantially, reducing the capacity to distort management priorities noted under targets – as such it is a natural development of the move from star ratings to the annual healthcheck.	Low

## Appendix three

### Information for patient choice and activation

Area	Data	Sources	Data type	Data Available/ Analysis undertaken
Clinical process	<ul style="list-style-type: none"> <li>% patients with BMI recorded</li> <li>% patients given smoking cessation</li> <li>% patients given HbA1c test</li> <li>% patients given retinal screening</li> <li>% patients with blood pressure tested</li> <li>% of patients with micro albuminuria testing</li> <li>% of patients with micro albuminuria treated with ACE inhibitors</li> <li>% of patients with measured cholesterol</li> </ul>	Quality and outcomes framework	Measure of process use relative measures to identify and judge on consistently outlier status	<p>All available at practice level</p> <p>Set of measures agreed through international consensus drawing on, for example, NQF in US (which has a smaller set for diabetes) and Structured Dialog from Germany (which will probably have some data on this area). Alternatively QOF weightings could be used although these are not particularly discriminatory.</p>
Patient process	<ul style="list-style-type: none"> <li>% of patients who knew their HbA1c value?</li> <li>% of patients who knew when to take medication</li> <li>% of patients who knew how much medication to take</li> <li>% of patients who received the right amount of verbal information about their diabetes</li> <li>% of patients who received the right amount of written information about their diabetes</li> </ul>	Healthcare Commission patient survey	Relative measures of variance	<p>All these data are currently available. Note this is currently at PCT level while QOF data is at practice level. We select five from approximately 80 indicators which are good sentinel measures.</p> <p>These were collected as a one off exercise and an 80 measure survey could not be repeated every year but collection of 10% of the data each year would be a justifiable exercise</p>
Clinical outcomes	<ul style="list-style-type: none"> <li>% of patients with last HbA1c less than 7.4 and between 7.4 and 10</li> <li>% of patients with blood pressure of 145/85 or less</li> <li>% of patients with cholesterol of 5 mmol/l or less 145/85 or less</li> </ul>	QOF	Outcome measures, judgement criteria as for clinical process above	All available at practice level
Patient outcomes	<ul style="list-style-type: none"> <li>% of patients admitted to hospital because of their diabetes</li> <li>% of patients saying that diabetes affects their day to day living</li> </ul>	HC patient survey	Relative measures of variance	All these data are currently available. Note this is currently at PCT level while QOF data is at practice level.



---

## Glossary

**Access:** availability of a required medical service. In the UK context this in nearly all cases refers to the time spent waiting to receive the service.

**Activation:** the patient's willingness to be involved and assertive in the decision-making in their own care.

**Case-mix:** the relative complexity of the cases being seen by a particular doctor, department or hospital.

**Clinical governance:** a system through which NHS organisations are accountable for continuously improving the quality of their services and safeguarding high standards of care, by creating an environment in which clinical excellence will flourish.

**Clinical practice guidelines:** evidence-based guidelines which cover the appropriate treatment and care of patients with specific disease and conditions.

**Compassion:** in our classification of measurement of quality this encompasses the treatment of patients as human beings, ensuring dignity and respect, concern for their suffering, and care taking place in decent conditions.

**Construct validity:** how well an indicator actually measures what it purports to.

**Consumer:** one who consumes healthcare. In this context more or less synonymous with patient, but implies a more active approach to seeking healthcare, and particularly a willingness to "shop around" between different healthcare providers.

**Core standards:** The 24 Core standards determined by the DH with which all NHS healthcare providers must comply.

These cover clinical and cost effectiveness, governance, patient focus, accessible and responsive care, care environment and amenities, Public Health.

**Domain:** these are the categories of judgement of care. We propose four for accountability purposes, safety, access, compassion and competence.

**Elective care:** non-emergency care, typically admissions into hospital from the waiting list.

**Gaming:** altering practice in order to give the impression of improved performance rather than actually improving performance. An example of gaming would be to redesignate a corridor as a ward in order to achieve an A&E throughput target.

**Indicator:** a measure of some aspect of care which allows a judgement to be made about performance. For example this might be a percentage of patients receiving care in accordance with guidelines, or could be the percentage of patients experiencing a good or poor outcome as a result of a particular procedure.

**Informatic(s):** collection, handling, analysis and presentation of information – covering all information related disciplines including information and communication technology, data management, analysis and statistics and data presentation.

**Institution:** healthcare providing or commissioning organisation.

**Measure:** a numeric summary of the general experience of, in this instance, healthcare.

**National Clinical Audit:** clinical audit is a quality improvement process that seeks to improve patient care and outcomes through systematic review of care against

explicit criteria. National Clinical audits are schemes carried out nationally to uniform standards, allowing national comparison of performance.

**Outcome - clinical perspective:** hard measurements of success or failure which include mortality and hospital readmission rates.

**Outcome - patient perspective:** the patient's perception of whether their state of health has improved following discharge from hospital.

**Outmigration:** patients being forced to go outside of their local area for their treatment, particularly because their local hospitals are unwilling to treat them.

**Outlier:** a data point so extreme as to be statistically significantly different from the average. The data point may refer to an organisation or an individual. For example, Dr Shipman was an outlier among GPs for mortality of patients.

**Patient:** a person requiring, undergoing, or seeking healthcare.

**Performance management:** the system of hierarchical accountability and control in all private and public sector organisations. This typically relates to the achievement of set objectives which may or may not be subject to formal incentives.

**Process - clinical perspective:** measures of how well the correct clinical procedures were followed such as following NSF and NICE guidelines.

**Process - patient perspective:** measures of the patients experience of the process of care. May include being treated with dignity and respect, cleanliness and quality of food and accommodation.

**Provider:** an organisation providing healthcare. Typically an NHS trust, independent sector hospital, or treatment centre.

**Purchaser:** an organisation purchasing healthcare: Primary Care Trusts in the UK, private sector insurance companies in the US.

**Quality and outcomes framework (QOF):** nearly 150 indicators of the quality of management and care of patients with a range of diseases used to assess performance of GP practices and support the new GP contract.

**Safety:** ensuring that care will “first do no harm”, that known risks are minimised and approved safety procedures are followed.

**Sentinel measures:** these are measures which identify something so unusual, or where the message sent is so unequivocal, that further investigation is required. For example, mortality following a day case admission.

**Specialty:** particular branch of the medical profession e.g. surgery.

**Societal value:** something which has importance for society as a whole, rather than just the individual patient. Often this can be linked to a political manifesto, e.g. Labour's pledge to reduce waiting times for non-emergency admissions represented a popular desire for these to be reduced, and thus reducing waiting times may have a societal value.

**Structure:** the adequacy of facilities and equipment, the qualifications of medical staff, administrative structure.

**Sub-specialty:** sub branch of a medical 'specialty' (see above) e.g. heart surgery.







Modern enterprises see information on their performance as vital – what is extraordinary about the NHS is that it spends so much on hospital care and knows so little about what this spending achieves. Outcome data currently focuses on mortality and readmission. This excludes around 90 per cent of hospital admissions. Whenever information has been collected on healthcare, it has revealed serious failings that require corrective action, and has identified high performers from whom others can learn.

This report examines measurement schemes in the UK and US that provide essential lessons for policy makers and incorporates groundbreaking polling that reveals how patients would react to information on quality. The research shows that measures of quality must have both visibility and credibility for clinicians and that publishing measures, rather than just reporting confidentially back to providers, increases the likelihood of driving up quality. The authors conclude that the effects of linking formal incentives (payment) to quality are ambiguous and that although patients want quality of care information, it is not clear that patient choice is stimulated by the publication of performance data.



£10.00

ISBN: 978-1-906097-07-3

Policy Exchange  
Clutha House  
10 Storey's Gate  
London SW1P 3AY