



USES & ABUSES OF PERFORMANCE DATA IN HEALTHCARE.

Authors:

Joanne Shaw

Joanne is a non-executive director and board chair with a combined private, public and third sector background. She is a member of the National Audit Office board. Until December 2013 she was chair of NHS Direct. She has a strong interest in the use of new communication channels for health and medicines. In her professional roles and in her writing for health publications she is known for advocating partnership between patients and health professionals and supporting people to make better informed choices about their health.

Roger Taylor

Roger Taylor is a consultant at Dr Foster where he is responsible for public information, including the Dr Foster Hospital Guide. Roger was a co-founder of Dr Foster and has been a strong advocate for the role of greater transparency as a mechanism for consumer protection as well as a driver for quality improvement in public services.

Katy Dix

Katy Dix initially worked for Andersen Consulting in their strategy consulting group before specialising in healthcare projects at Deloitte Consulting. She holds a Masters in Health Administration (MHA) and has worked in healthcare for over 15 years in both Australia and the UK. Her clients have included large private hospitals, pharmaceutical companies and health authorities, with an emphasis on leading focus groups, conducting literature reviews and writing structured reports.

Contents

1. Executive summary	3
2. Foreword	7
3. How this report was produced.....	9
4. Scope	10
5. Positive effects of performance data ('uses')	13
5.1. Targets	14
5.2. Pay for performance	14
5.3. Publication	15
5.4. Professionally-led improvement.....	16
6. Adverse effects of collecting performance data ('abuses')	19
6.1. Tunnel vision	19
6.2. Adverse selection	20
6.3. Erosion of professional motivation	21
6.4. Ceiling effect	22
6.5. Bullying/intimidation.....	23
6.6. Gaming.....	24
Gaming the data	24
Distorting the process.....	26
6.7. Distraction.....	27
6.8. Other public services	28
7. Discussion	30
8. Recommendations.....	34
8.1. Make data quality as important as hitting targets	34
8.2. Measure the context not just the indicator	35
8.3. Avoid thresholds	36
8.4. Be open	36
8.5. Apply measures fairly	37
9. Acknowledgement and thanks	38

1. Executive summary

Hopes of improving healthcare through better measurement and the use of information in healthcare management are being undermined by weaknesses in the generation and use of data and metrics. This report outlines five steps that could address these problems.

Measurement of performance is essential for transparency and accountability, and to support improvement. To do this effectively it must reconcile important and potentially conflicting goals:

- Meaningful accountability, requiring real consequences from underperformance
- An environment that encourages open and honest reporting

Accountability regimes should acknowledge and address in their design the fact that those holding NHS organisations to account and those being held to account have an interest in accepting apparent improvements in performance and not questioning too deeply whether the results flow from changes in data recording or superficial changes in the way care is delivered.

This report aims to improve the way data and measurement are used, by looking at the creation and use of performance indicators within the broader canvas of today's health system and people management:

- Recognising practical issues, constraints and pressures
- Learning from academic research, expert opinion and new analysis
- Offering constructive suggestions for clinicians, managers, policy makers, patient groups and elected representatives
- Supporting local leaders as they take greater responsibility for health and social care systems

There are impressive examples of targets, measurement and reporting leading to significant improvement, including:

- Dramatic reductions in English patient waiting times
- Lower mortality for cardiac surgery in Great Britain and Ireland
- Improvements in prostate cancer outcomes in Germany
- Better cardiothoracic care in Leicester

Performance measurement can have a range of unintended adverse consequences:

- **Tunnel vision** – focusing on aspects of clinical performance that are measured and neglecting unmeasured areas
- **Adverse selection/ inequity** – avoiding the most severely ill patients or excluding disadvantaged groups
- **Bullying** – intimidating staff to achieve performance targets or to adjust data
- **Erosion** – diminution of intrinsic professional motivation as a key driver of high-quality healthcare
- **Ceiling effect** – removing incentives for further improvement and potentially influencing top performers to reduce quality
- **Gaming** – distorting the process of care in order to meet targets or manipulating data to misrepresent actual performance
- **Distraction** – challenging, obfuscating or denying data which suggests under-performance instead of fixing performance problems

Examples of adverse impacts of measures in healthcare include:

- Areas with high GP Quality and Outcomes Framework (QOF) scores for diabetes care processes also having high rates of diabetes-related amputations (outcomes)
- Cardiac surgery report cards leading to selection behaviour by US providers and worse outcomes for patients
- Bullying culture within the NHS
- Gaming waiting time and mortality data in English hospitals
- Distortion of patient pathways to meet cancer treatment targets
- Arguments about data quality diverting attention from poor care

The reference group brought together to discuss these issues identified five key aspects of legitimate performance measurement:

1. Outcome focussed: outcomes count for more than processes
2. Patient focussed: measures should capture the patient's experience and outcome
3. Clinically credible: clinicians should have a say over how they are measured
4. Based on local need: local health economies should have flexibility in how nationally mandated goals are achieved and room to set local goals

5. Based on the performance of the whole system, not just part of it

A number of policies have attempted to address these issues but have foundered for a variety of reasons including:

- Fear of visible short-term performance decline in response to long-term outcome targets
- Variable data quality
- Weak information and analytical capabilities at health economy level
- Barriers to sharing patient data
- IT limitations preventing whole system measurement

We have identified five specific steps that are not currently being addressed by policy, which could significantly reduce data abuse and increase the benefit that could be gained through use of performance data.

1. Make data quality as important as hitting targets

Failure to tackle data quality risks undermining the entire enterprise of performance management. The risks range from failure to tackle misreporting and incomplete or inaccurate data recording, to not maintaining or improving data sources to support the range of metrics needed. There should be a long-term audit programme to assure data integrity. Where possible, data should be drawn from information sources shared with patients and used in other contexts, to support consistency and accuracy.

2. Measure the context not just the indicator

The negative consequences of many aspects of performance management are widely known within the health system but less often publicly acknowledged. Performance measures need to be monitored along with sets of counterbalancing metrics that can identify issues such as pathway distortion. Measures need to be constantly monitored and reviewed in the light of experience. This task could be undertaken by multi-disciplinary specialist groups, including royal colleges and patient organisations.

3. Avoid thresholds and consider the potential to incentivise gaming in the design of metrics

Performance management metrics are often designed with a view to simplicity of implementation and communication, and with insufficient thought to the likely negative consequences. Performance measures should be assessed against these risks. Thresholds should be avoided wherever possible.

4. Be more open

Make data underlying performance management widely available and promote ongoing assessment of the degree to which metrics are being gamed. Build in regular reviews of performance management regimes. Expect to make annual refinements to the way in which performance is measured to minimise fine tuning of systems to meet targets rather than achieving the desired benefit for patients.

5. Apply measures fairly

Performance management in the NHS has been seen as a form of arbitrary justice which fails to recognise legitimate mitigating factors such as resources and pressures outside the control of the organisation. The most serious issue is the creation of conflicts between the requirements of performance management and the needs of individual patients. Performance management regimes must explicitly allow for the possibility of breaches in patient interests.

Improving the way in which performance of healthcare is measured and the way in which that information is used are central to the improvement of healthcare itself. We hope that following the general election, whoever is in power, this issue will be given the appropriate level of priority in determining the future management of the NHS.

2. Foreword

In 2003 the Audit Commission published a paper pointing to widespread inaccuracy in reported patient waiting times and highlighting deliberate manipulation of data in three of the 41 NHS trusts it examined. The Commission also uncovered practices designed to improve the appearance of waiting time performance that were bad for patient care, such as offering appointments at short notice and re-setting the clock if patients could not attend. In January 2014, the National Audit Office again reported problems with the completeness, consistency and accuracy of patient waiting time data. Once again three (different) hospitals were singled out for mis-recording waiting times for treatment or diagnostic investigations. Two papers, ten years apart, same problems.

Meanwhile in Mid-Staffordshire, an excessive focus on financial and performance targets led Robert Francis to observe in 2013, a 'culture focused on doing the system's business – not that of the patients' and which 'ascribed more weight to positive information about the service than to information capable of implying cause for concern.'¹ In 2015, Bill Kirkup, investigating avoidable patient deaths in Morecambe Bay NHS Trust, found that 'performance and management targets [were] given clear priority. Clinical issues were not addressed.'²

Francis and others have argued that a complete culture change is needed to fix the difficulties afflicting the NHS. Others attach blame more narrowly, to an obsession with measurement. Top-down, central targets will never work, they say, and should be abolished, to be replaced by local accountability and locally-driven improvement initiatives. Process measures should be abandoned altogether and replaced by outcomes. The correct response to under-performing hospitals should be more resources and support, not sanctions.

This leaves our health system with what looks like an intractable dilemma. Measurement is essential for improvement. Standard ways of measuring and reporting performance are needed so patients and taxpayers can know what is being delivered to them, in their name, and with their money. Clinicians need measures to understand what's working, to compare their performance with others, and to help identify opportunities for improvement. Elected representatives need measures to explain to the public what they are aiming to achieve and to demonstrate how the NHS is performing under their stewardship so that the electorate can hold them to account. Without meaningful consequences from poor performance within the healthcare system, or worse,

¹ 'Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry', Francis, Feb 2013

² 'The Report of the Morecambe Bay Investigation,' Kirkup, March 2015

where poor performance is rewarded with more resources, genuine accountability is impossible. But people are people, and when targets are mandated centrally, with carrots (money) and sticks (various and unpleasant), gaming is inevitable. Well-established change management experience and theory tells us that sustainable improvement requires engaging the hearts and minds of front-line teams. Effective change projects need the freedom to decide what to measure, as well as how to improve.

In a democratic society that aspires to improve the health and care of its population, measurement and transparency in healthcare cannot and should not go away. Quite the reverse – we need more of it, not less. But we also recognise the complications that come from human nature, and the inevitable temptation to want to make things seem better than they really are. Our aim in this paper is to reflect the real world, recognising the tensions and offering practical suggestions about how data and measurement can be used to better effect, so making some of the most pernicious problems from the abuse of data less prevalent in the future.

It comes at a time when moves towards greater local integration of health and social care are likely to lead to more involvement of locally elected representatives, irrespective of the outcome of the general election. The recommendations in this paper should help those elected leaders set and use local targets, objectives and measures in their own health and social care systems, avoiding some of the problems seen in past national target-setting. We hope that the report will support new partnerships of local integrated care commissioners in their efforts to use data well.

3. How this report was produced

An initial literature review was conducted to derive a taxonomy for understanding the ways in which performance data can be abused, and to collect examples from academic literature and the media of misuse of performance information.

This framework was shared with the group of clinicians, managers, data experts and commissioners whose names are listed at the back of this report. The group met at a round table event in January 2015 to refine the taxonomy and to share further positive and negative examples of the ways in which data was used effectively or ineffectively.

Hypotheses derived from the discussion were tested by reference to published data, and case studies developed to illustrate the issues that the group felt were most important. The final document benefited from further input by members of the reference group.

The views expressed in this report are those of the authors and not the views of either the contributors to the roundtable or the Dr Foster Ethics Committee.

4. Scope

Most people involved in healthcare would agree that better information has helped make the NHS in England safer, more efficient and more transparent, and will continue to do so in the future. At the same time, data collection is a significant burden on healthcare organisations and the abuse of data as a management tool has had highly dysfunctional effects for patients and healthcare workers alike.

A vast amount of information about health services is now collected, analysed and published, including activity levels, process metrics, clinical outcomes and patient experience. The results are made available through a whole range of channels including:

- Care Quality Commission risk ratings and assessments
- NHS Choices
- Dr Foster Hospital Guide
- Clinical audits
- NHS England dashboards (safety culture, Friends & Family scores)
- *Which?* maternity services reports
- The media and material gathered under freedom of information legislation.

Good data can spotlight excellent practice and illuminate dark corners where things are going wrong. Conversely, measurement, target setting and publication of results can become oppressive; activity can be distorted to produce more acceptable numbers; and arguments about data validity can distract attention from real issues, diverting scarce resources from much-needed improvement. We therefore decided to look at the uses and abuses of performance data, including measures of clinical quality, safety and efficiency. The aim was to draw on the real experience of healthcare practitioners in the UK and to look at international examples, in order to make recommendations on how to combat the main types of 'data abuse' and enable more constructive use of data and measurement.

This piece of analysis has focussed on the impact of performance measures on the health system, rather than on patient choice and behaviour, which may form the basis of a further project in the future. It is not intended to be a comprehensive review. The aim is to share ideas about how to encourage positive behaviours and reduce negatives in terms of the way:

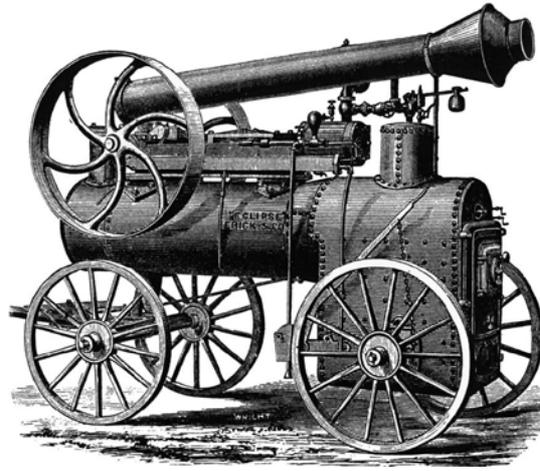
- Data is recorded, created, validated and processed
- Quality is measured

- Information is disseminated
- Data is used to hold individuals and organisations to account – both in private and in public

When organisations are held accountable by external bodies or the public for performance against targets, and particularly when financial incentives are attached, it is tempting for individuals and organisations to manipulate the data in ways that make their performance appear to meet the required standards. Indeed, some would say that a degree of ‘gaming’ the data is almost inevitable. On the other hand, measurement without any consequences flowing from underperformance, whilst less likely to lead to data abuse, is unlikely to satisfy voters or their representatives, or to lead to meaningful improvement.

The challenge is to use performance data to provide accountability and stimulate improvement, without leading to adverse effects that swamp the intended gains. This report starts with examples from the UK and overseas illustrating how improvement can follow from analysis and publication of performance data. It will then unpack some of the ways in which data has been abused within the NHS and other public services. It will discuss the learning from these negative and positive examples. Finally, it will make recommendations for a range of participants in the wider health system to encourage more constructive use of performance information in health.

Measuring performance can be like running a steam engine



Source: www.farmcollector.com

The earliest steam engines had a power efficiency of 0.5% with most of the energy generated from burning coal lost through poor design. James Watt's designs increased efficiency fivefold. At their peak the best engines were 100 times as efficient as those first developed.

To run a steam engine we shovel in coal, with the aim of maximising the power output. But as the engine builds power, it also generates waste steam and a horrible noise. The challenge is how to maximise useful power output whilst minimising lost energy. In the same way, we are looking for ways to gather and use performance data productively and minimise the amount of wasted heat and noise.

One way to make steam engines more efficient is to make them less powerful. In the same way, many people argue that the only way to make performance information more useful is by putting less stress on it and asking it to do less. Although that might increase efficiency, it reduces the overall work that can be done. Our objective is to find ways of using data that are both *efficient* and *powerful*.

5. Positive effects of performance data ('uses')

There are several possible routes by which collecting and sharing performance information can lead to improvements in healthcare. Clinical teams rely on data to benchmark their performance and understand what works. By showing the range of results that are achievable, valid data can highlight excellence and enable identification and sharing of good practice, thereby driving up standards across the board. Institutions and teams who can demonstrate excellence should attract recruits, patients and, in a system where resources follow the patient, transparent performance measurement should enable the best centres to grow at the expense of less-effective ones unless they are able to improve.

Published data can focus attention where performance is sub-standard and hasten the progressive elimination of poor and outdated practice. Comparative information can engage professional pride and act as a spur to individual or team competition to raise standards and improve results. It should deter organisations from offering services that they are not well equipped to provide at a high standard, and encourage individual practitioners to stop carrying out procedures where their results are below the average of their peers. All of these effects have been observed.

There are, however, relatively few unambiguous examples of positive change in health brought about through measurement and publication of data. Attribution of cause and effect is complicated by the complexity of the health system, political turbulence, the timescales needed to demonstrate improvement, and the multiplicity of other factors that affect health outcomes. To make sense of this complex landscape, we have defined four archetypal ways in which data is used to bring about change:

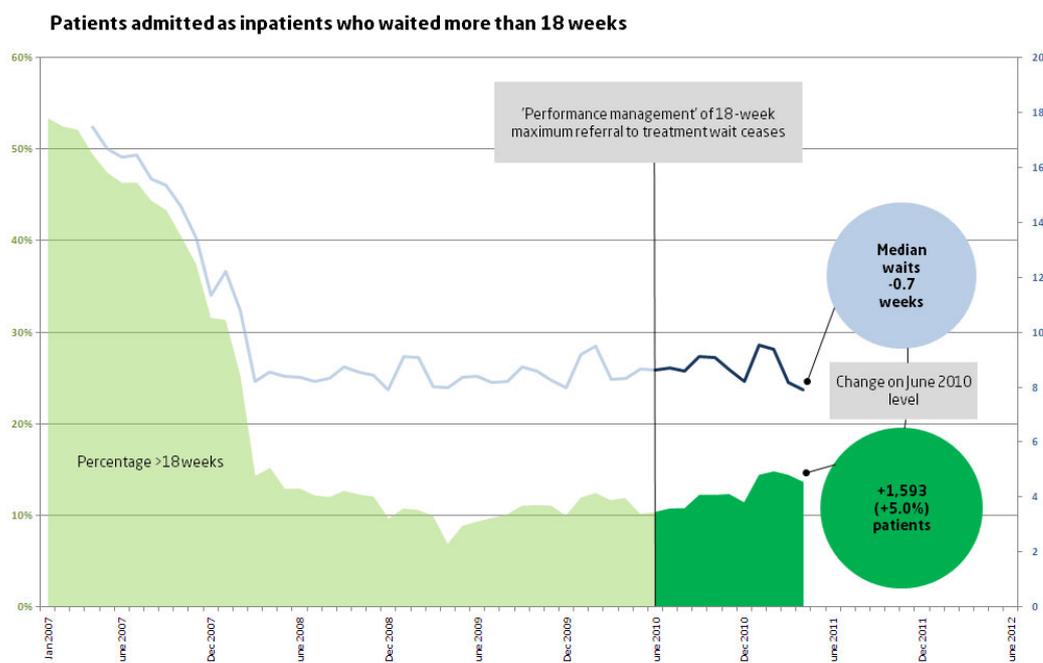
- **Targets** – specific goals with top down sanctions for under performance
- **Pay-for-performance** – financial rewards linked to measured performance
- **Publication** – putting information in the public domain
- **Professionally-led improvement** – using information to inform changes in practice in the absence of targets or other incentives

These aspects are often combined. Each is described overlaid in more detail, with examples that have been cited as evidence of their effectiveness and ability to bring about positive change.

5.1. TARGETS

The preoccupation with waiting times has resulted in some unintended effects in terms of patient pathways being distorted and data massaged to meet targets, but it has clearly had dramatic positive effects on patient experience.

Figure 1 – Patients admitted as inpatients who waited more than 18 weeks June 2007 - June 2012



Source: 'Hospital Waiting Times Tracker: April 2011', The Kings Fund, April 2011

Despite the fact that waiting times are only one component of the total patient journey, they are clearly important to the general public and hence to their elected representatives, although public perception of waiting times do not always mirror actual system performance³. Waiting times have also been shown to have a direct impact on outcomes⁴.

5.2. PAY FOR PERFORMANCE

The QOF reward and incentive programme governing the performance management and payment of general practitioners (GPs) was introduced as part of the GP contract in 2004. QOF awards surgeries points for:

- managing some of the most common chronic diseases including asthma and diabetes
- implementing preventative measures such as regular blood pressure checks

³'Understanding public and patient attitudes to the NHS', Healthcare Commission, Aug 2006

⁴'Impact of waiting time on the quality of life of patients awaiting coronary artery bypass grafting', Sampalis et al, *CMAJ* Aug 2001

- extra services such as child healthcare and maternity
- quality and productivity, including the avoidance of emergency hospital admissions
- minimum length of GP appointments.

Achievement points are converted into monetary rewards for practices. Introduction of the QOF had a major impact on activity within general practice and its effects have been extensively studied. The new GP contract as a whole cost £1.76 billion more than the government had expected,⁵ mainly because GPs had been expected to achieve 75% of the available points in the first year and actually achieved 90%. A 2011 review in the *British Medical Journal* (BMJ) concluded that there were substantial improvements in quality for all indicators between 2001 and 2007. Improvements associated with financial incentives seem to have been achieved at the expense of small detrimental effects on aspects of care that were not incentivised.⁶

5.3. PUBLICATION

In the USA, publishing the performance of hospitals, health professionals and providers is believed to drive quality improvement by changing the behaviour of health consumers or purchasers. Better-informed individuals can choose the health provider that they see as providing the best quality and value. As a result, demand for the poorer-performing providers should reduce.

Several studies in the US have examined the effects of the 1990 introduction of the New York State Cardiac Surgery Reporting System, which reports hospital mortality following coronary artery bypass graft (CABG) surgery. Many show a significant reduction in risk-adjusted mortality^{7,8} as a result of the system. There has been debate about the exact cause of this decrease in mortality, however, with some pointing out that cardiac surgery mortality rates had been decreasing for years. The challenge has been attributing causality to the introduction of public reporting.

In the UK, the publication of clinical outcomes has been less of a guide for health consumers and more of an aid to professional development. The Society for Cardiothoracic Surgery in Great Britain & Ireland has pioneered collection of surgical information in the UK, having run a register of all major heart operations since 1977.

⁵ 'Do GPs deserve their recent pay rise?', Timmins N., *BMJ*, Oct 2005

⁶ 'Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework', Doran T. et al, *BMJ*, June 2011

⁷ 'New York State's cardiac surgery reporting system: four years later', Hannan et al, *Annals of Thoracic Surgery*, Dec 1994

⁸ 'How a New York cardiac surgery program uses outcomes data', Dzuiban et al, *Annals of Thoracic Surgery*, Dec 1994

Society for Cardiothoracic Surgery in Great Britain & Ireland National Adult Cardiac Surgery Audit⁹

The National Adult Cardiac Surgery Audit has collected information about all major heart operations carried out in the UK since 1977. The audit allows comparison of clinical performance with national and international standards, and provides useful data on changing trends within the specialty. Data down to individual surgeon level has been published since 2005.

According to the published clinical outcome data, there has been a 25% reduction in mortality for adult cardiac surgery since 2003 and risk-adjusted mortality for cardiac surgery is 25% below the European average.

There is little or no evidence that publishing the data has had adverse effects, such as avoidance of high-risk patients.

5.4. PROFESSIONALLY-LED IMPROVEMENT

Outcome measurement focuses attention on the factors that are most important to clinicians and patients and are less susceptible to gaming and distortion. The Martini Klinik in Germany has been recording and using health outcome data for many years.¹⁰ As the largest prostate cancer treatment centre in the world, it reports a high level of patient satisfaction and a lower-than-average complication rate¹¹ from surgery. It has achieved this by focussing on outcomes, in the following ways:

- Since 1994 the Martini Klinik has collected clinical, administrative and some outcome data for every patient treated. Its post-discharge survey achieves a 90% response rate.
- Clinicians receive data on their own outcomes and those of their colleagues six-monthly. Reports include basic information (case volume per surgeon), data related to surgery (like average blood loss) together with outcome data about patients' experience in the months after surgery. Outcome data for the previous six months is compared with the results of earlier years.
- Regular presentations are made, to showcase recent literature on advances in prostate cancer care. Further meetings are led by clinicians, to discuss their outcomes with visiting doctors.

⁹ Society for cardiothoracic surgery in Great Britain & Ireland, <http://www.scts.org/>

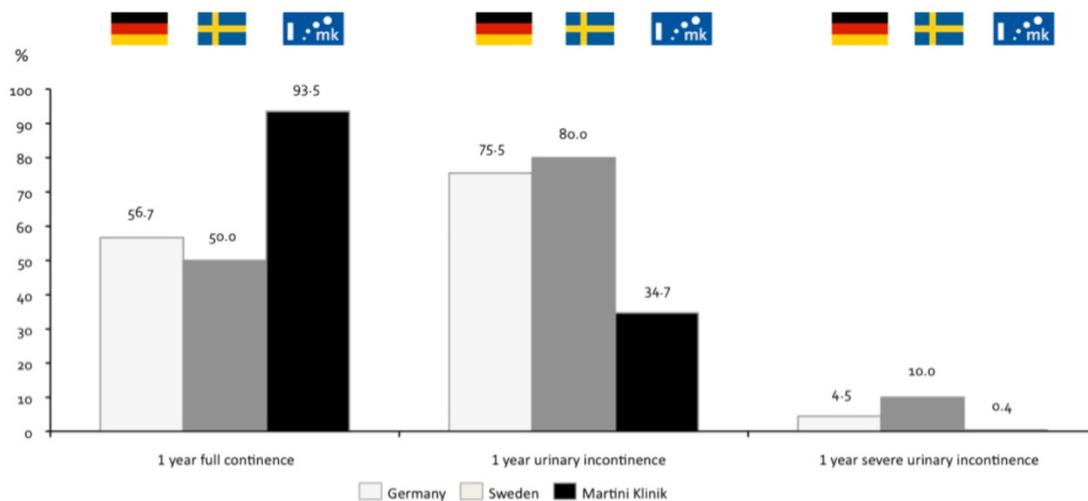
¹⁰ 'Martini Klinik: Prostate cancer care, case study', Porter et al, *Harvard Business Review*, June 2014

¹¹ 'The average one-year continence rate at Martini was 93.5% versus 56.7% in Germany; 4.5% of public health plan patients complained of severe urinary incontinence compared to 0.4% of Martini Klinik patients. In 2012, 34.7% of Martini Klinik patients reported severe erectile dysfunction one year after operation compared to 75.5% for average German public health plan patients.' Source: BARMER GEK Krankenhaus Bericht 2012, Martini Clinic Database

- There is a half-yearly review meeting to analyse outcomes at the individual surgeon level. Risk-adjusted outcome measures are compared in detail, taking account of surgical technique (eg, nerve-sparing surgery, robotic surgery) and controlling for the assisting surgeon.
- Surgeons with higher than expected complication rates are asked to assist in operations with more experienced surgeons. They are also observed by the higher-achieving surgeons in their subsequent operations.

Clinicians from the Martini Klinik have since worked with the International Consortium for Health Outcomes Measurement's (ICHOM) localised prostate cancer working group to create a global set of outcomes that matter to patients so that they can be benchmarked internationally.

Figure 2 - Martini Klinik: outcome performance vs benchmarks



Source: ICHOM

Measuring outcomes has also had a positive effect in Sweden, where quality registries analyse long-term outcomes for patients with similar conditions, or who have undergone the same treatment. The Swedish Total Hip Replacement Registry, for example, was started in 1979 to record the outcome of primary hip replacements and to provide information regarding serious complications. Initially demographic data, diagnosis, surgical technique and type of implant were recorded and, in 2002, patient-reported outcome such as pain relief, satisfaction and gain in health-related quality of life were added. Analysis of this data has allowed doctors to compare the long-term performance of both procedures and implants. Sweden now has one of the world's lowest failure rates for artificial hips, with the revision burden falling from 17% to 7% over the lifetime of the registry.

Significant progress is being made in Leicester through improvements in care prompted by analysis of mortality data. In 2012 life expectancy in Leicester City was significantly worse than England and the East Midlands, and the life expectancy gap with England was widening. The main factors contributing to the widening gap were mortality from cardiovascular disease (CVD) and respiratory diseases. Ten years earlier, premature CVD mortality in Leicester was 27% higher than England as a whole, but by 2012 excess CVD mortality had risen to 53%.

Further interrogation of morbidity, diagnosis and treatment data showed that only 50% of relevant patients were being treated with anticoagulation therapy. Under a carefully managed step-by-step development programme led by Dr Umesh Roy, 55 local practices were engaged, trained and supported to take on responsibility for initiating, monitoring and managing anticoagulation therapy. Over a 12-month period the proportion of relevant patients on therapy rose from 50% to 70%, meaning that an additional 800 people were being treated and 300 new diagnoses were made. Validated models suggest that this will result in a 65% reduction in the incidence of strokes in that group.

These examples have been chosen to illustrate how the analysis and publication of comparative performance data can be the starting point for initiatives that led to meaningful improvements for patients, through targets, publication of results, payment for performance and clinician-led improvement programmes. The UK waiting time initiative differs from the others in its focus on a single process measure. In general, though, greater focus on outcome measures makes it easier to engage professionals and harder to distort the process or the data to create a misleading impression of performance. This can be seen more clearly through some examples of data abuse in the following section.

6. Adverse effects of collecting performance data ('abuses')

The previous section has provided some evidence that the theoretical value of performance information can deliver real benefits in practice. But unintended negative consequences of collecting performance data in healthcare are very striking. The examples in this section have been categorised starting from a 2014 paper by Mannion in the *Journal of Health Policy Management*¹² and developed in discussion with our reference group of practitioners, managers and data specialists.

- **Tunnel vision** – focusing on aspects of clinical performance that are measured and neglecting unmeasured areas
- **Adverse selection/ inequity** – avoiding the most severely ill patients or excluding disadvantaged groups
- **Bullying** – intimidating staff to achieve performance targets or to adjust data
- **Erosion** – diminution of intrinsic professional motivation as a key driver of high quality healthcare
- **Ceiling effect** – removing incentives for further improvement, and potentially influencing top performers to reduce quality
- **Gaming** – distorting the process of care in order to meet targets, or manipulating data to misrepresent actual performance
- **Distraction** – challenging, obfuscating or denying data which suggests under-performance instead of fixing performance problems

The continuum covered by these behaviours extends from mildly discreditable, through clearly unethical, to criminal. The following examples have been chosen to illustrate the categories without necessarily implying whereabouts on that spectrum they lie. Other public services outside healthcare also demonstrate these problems.

6.1. TUNNEL VISION

The old adage that 'what gets measured gets managed' can lead to an excessive focus on aspects of clinical performance where performance data is collected. Tunnel vision can happen where organisational units are responsible for component parts of a larger process and focus exclusively on measures of their

¹² 'Take the money and run: the challenges of designing and evaluating financial incentives in healthcare; Comment on "Paying for performance in healthcare organisations"', Mannion, *International Journal of Health Policy Management*, Feb 2014

own performance while neglecting the overall outcome. This can be exacerbated where financial reward is attached to particular aspects of performance.

Because processes are easier to measure unequivocally than outcomes, and can demonstrate improvement more quickly, performance management regimes and incentive regimes may put greater weight on the process than the outcome. Tunnel vision can then result in improvements in processes though they yield no benefit to patients.

In April 2014 Diabetes UK noted that geographical variation in the rate of diabetes-related amputations was getting worse. They observed that people with diabetes in the worst-performing area were seven times more likely to have an amputation than people in the best-performing area.¹³ Surprisingly, the worst area for amputations achieves a higher percentage of QOF points for diabetes than the UK average.¹⁴ It therefore meets the requirements of the QOF targets which are intended to show how effectively GPs care for people with diabetes, but outcomes remain poor for many of its diabetic patients.

Studies have shown that performance on QOF metrics in heart disease are not associated with improved outcomes for patients.

Tunnel vision can be difficult to evidence because, by definition, the impact of concentrating on the aspects of care which are measured is felt in the areas which are effectively invisible. However, an analysis of the quality of primary care after the introduction of QOF targets showed that:¹⁵

'Against a background of increases in the quality of care before the pay-for-performance scheme was introduced, the scheme accelerated improvements in quality for two of three chronic conditions in the short term. However, once targets were reached, the improvement in the quality of care for patients with these conditions slowed, and the quality of care declined for two conditions that had not been linked to incentives.'

6.2. ADVERSE SELECTION

Performance measures in healthcare generate a potential incentive to avoid the most severely ill patients or exclude disadvantaged groups. The British Medical Association (BMA) has said: 'Some surgeons are deterred from taking on very complex, high-risk procedures because published simplistic league tables count against them.'¹⁶ Consumers and patients have much less information about treatments, which creates the potential for moral hazard and adverse selection.

¹³ 'Amputation postcode lottery getting worse', Diabetes UK, April 2014

¹⁴ QOF data for 2012/ 2013, NHS Networks

¹⁵ 'Effects of pay for performance on the quality of primary care in England', Campbell et al, *New England Journal of Medicine*, July 2009

¹⁶ 'BMA warns against simplistic surgeon tables', BMA, December 2012

Given the lack of published outcomes data on doctor/surgeon performance to date in the UK, most evidence for this comes from the US. Early surveys certainly showed that initial publication of cardiac surgeon performance led to a significant proportion stating that they were less willing to operate¹⁷ or perform angioplasty¹⁸ on severely ill patients. A further study into the effects of introducing healthcare report cards in the US¹⁹ showed,

'Healthcare report cards may address important information asymmetries in markets for healthcare, but they may also give doctors and hospitals incentives to decline to treat more difficult, severely ill patients. ...Using national data on Medicare patients at risk for cardiac surgery..., cardiac surgery report cards in New York and Pennsylvania led ...to higher levels of resource use and to worse health outcomes, particularly for sicker patients.'

Evidence for this effect is contradictory, with others saying that systematic bias against high-risk patients in coronary artery bypass surgery has not clearly been proved²⁰ and we have yet to see whether in the UK 'giving the public access to surgeon-specific mortality data shifts the emphasis from "patient care to self-preservation"' as predicted by an editorial in the *BMJ*.²¹ More recently, the secretary of the Society of Cardio Thoracic Surgeons (SCTS) has said that there is no hard evidence that surgeons are avoiding operating on sicker people with a higher risk of dying on the operating table – and, on average, surgeons have seen the risk level of patients operated on rise in the period since the SCTS began public reporting. However a recent survey has found that 25% of heart surgeons say they are now less likely to take such cases and 75% claim to have seen risk-averse behaviour in a colleague²².

6.3. EROSION OF PROFESSIONAL MOTIVATION

Excessive focus on meeting performance targets can potentially lead to an erosion of professional motivation. This can take two forms. First it may lead to a 'work to rule' problem, whereby the essential goodwill of individuals in the health system is lost. The NHS relies upon the professional desire of its employees to do the right thing by their patients. The NHS benefits from countless hours of unpaid work, put in by staff in the interests of getting the job done. The pressure to measure every step of the process may mean that this goodwill is lost and patients will suffer as a result. Performance measures can never take into

¹⁷ 'Influence of cardiac surgery performance on referral practices and access to care', Schneider & Epstein, *New England Journal of Medicine*, July 1996

¹⁸ 'The influence of public reporting of outcome data on medical decision making by physicians', Narins et al, *Archives of Internal Medicine*, Jan 2005

¹⁹ 'Do report cards tell consumers anything they don't already know? The case of Medicare HMOs', Dafny & Dranove, *Rand Journal of Economics*, Autumn 2008

²⁰ 'Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients?' Hannan et al, *Medical Care*, Jan 1997

²¹ Surgeons may turn away high-risk patients over death rates, consultants warn', *The Telegraph*, August 2014

²² Ibid.

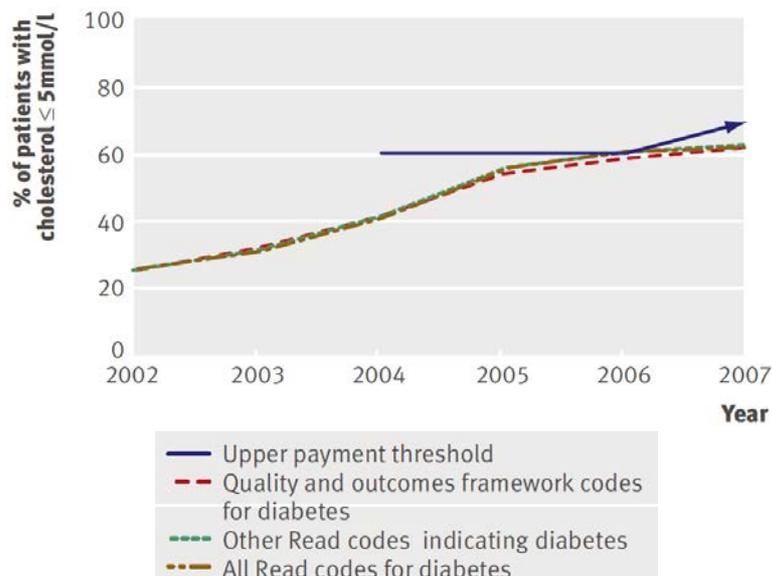
account every patient's circumstances and any metric risks creating situations where the requirement of performance management conflicts with the interests of the patient in front of the doctor. At worst, professionals may conclude that the only way to progress in their career is to conform with the performance management regime even where it conflicts with their personal values.

6.4. CEILING EFFECT

The ceiling effect is a recurring theme in the context of target approaches to public services, whereby a uniform output target applied to all units in a system removes the incentive for excellence, and may have the opposite effect of influencing top performers to reduce their achievements²³.

Studies of the management of diabetes between 2002 and 2007²⁴ investigated whether the introduction of the QOF had any effect on diabetes care. They found consistent improvements in all diabetes indicators over the five-year period but a slowing down after 2005. This attenuation may be explained by the lack of further incentive once the upper QOF payment threshold has been reached.^{25,26} Figure 3 illustrates the improving trend to 2005, plateauing in later years once the QOF target was reached.

Figure 3 - Effect of QOF on diabetes care: cholesterol levels for type 1 diabetics



Source: Dr Foster

²³ 'Gaming in targetworld: The targets approach to managing british public services', Hood C., *Public Administration Review*, July/Aug 2006

²⁴ 'Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study', Calvert M. et al, *BMJ*, May 2009

²⁵ 'The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review.', Langdown & Peckham, *Journal of Public Health*, June 2014

²⁶ 'Effects of pay for performance on the quality of primary care in England', Campbell S et al, *New England Journal of Medicine*, July 2009

6.5. BULLYING/INTIMIDATION

The pressure to record performance data that meets requirements can lead to bullying or intimidation of healthcare workers, both from within and from outside an organisation. The highest profile example of bullying and intimidation in recent years took place in Mid Staffordshire. But wider concerns about a bullying culture and the treatment of whistleblowers throughout the NHS led to the 2015 'Freedom to Speak Up'²⁷ report, also by Sir Robert Francis. It referred to the 2013 NHS Staff Survey, which revealed that nearly 30% of staff felt it would not be safe to raise a concern to management. One contributor to the review commented:

*"[The] NHS has a culture of bullying and harassment that means clinicians could not raise issues in clinical care and are pressured to put targets over ethics. If there is such a culture then it is because the majority of managers or clinicians in positions of authority are driving it/ managers recruited/ promoted to those positions because of their ability/ willingness to push this agenda."*²⁸

Following an allegation of manipulated waiting list data at NHS Lothian, an inquiry²⁹ including surveys and interviews with staff generated comments such as:

'Some senior managers bully us with constant targets, targets, targets, shouting and relentless pressure.'

'If you don't reach your targets you can collect your P45' (almost a standard phrase reported from a number of interviewees).

The report acknowledged a degree of truth in the claim that 'the Lothian Way is often referred to as "the bullying way"' and pointed to the 'context of extremely challenging times for NHS Lothian', with the hospital 'under significant pressure to deliver faster access to services for its patients whilst at the same time facing a challenging financial environment'.

Bullying can also be felt from outside the organisation. In November 2014, the chief executive of the Heart of England NHS Foundation Trust resigned, blaming the health regulator Monitor for contributing to a 'blame-based culture' and for using language that was:

*'Undermining leaders... which implies the leadership might only do the right thing if forced to do so. [Thus] we risk framing the inability to achieve targets as a behavioural issue.'*³⁰

²⁷ 'Freedom to speak up: An independent review into creating an open and honest reporting culture in the NHS', Francis, Feb 2015

²⁸ Ibid

²⁹ 'Investigation into management culture in NHS Lothian', Bowles & Associates, May 2012

³⁰ CEO Diary, 10 November 2014, Mark Newbold

The Trust had reported poor waiting time figures³¹. The outgoing CEO acknowledged the need for improvements, but questioned 'whether a more understanding and supportive approach would achieve quicker improvements for patients.'³²

Senior leaders acknowledge a deep-rooted problem of bullying culture within the NHS, which is often related to the pressure to demonstrate performance against targets. There are reports of staff being bullied to meet targets and bullied into manipulating data in order to make it seem that targets have been met. The recent 'Freedom to Speak Up' report shows that staff who report poor practice and falsification of data are themselves at risk of being bullied and intimidated.

6.6. GAMING

Excessive emphasis on performance data often leads to 'gaming'. This may involve gaming the data itself, for example, manipulating exclusions from waiting lists. It may also involve distorting the process of care, for example the practice of keeping patients waiting in ambulances to avoid the clock starting upon entry to the A&E department.

Gaming the data

Two audit reports into NHS waiting lists, ten years apart, show clear evidence of gaming waiting time data. In 2003 the Audit Commission conducted spot checks at 41 trusts and found evidence of deliberate misreporting of waiting list information at three of them. In a further 19 trusts, auditors found evidence of reporting errors in at least one waiting list performance indicator.³³

After the introduction of 18-week waiting time targets, a further study in 2014 by the National Audit Office³⁴ showed that there were still errors in trusts' recording of patient waiting times. They reviewed 650 orthopaedic patient waiting times across seven trusts and found that more than half were not supported by documented evidence or were incorrectly recorded. In addition, mis-recording of data was identified at three trusts, including one that had altered patient appointment and medical records on its cancer waiting times system.

Hospital mortality rates have the powerful advantage of measuring a health outcome that everyone cares about, but, as is well established, are themselves vulnerable to gaming. Hospital Standardised Mortality Ratios compare actual with expected death rates. Expected death rates are adjusted to reflect the number of patients admitted for palliative care, for whom death is the expected outcome. Therefore the proportion of patients coded as receiving palliative care has an

³¹ More than 5,500 patients had to wait more than the four-hour target at three of the Trust's hospitals in Birmingham and Solihull.

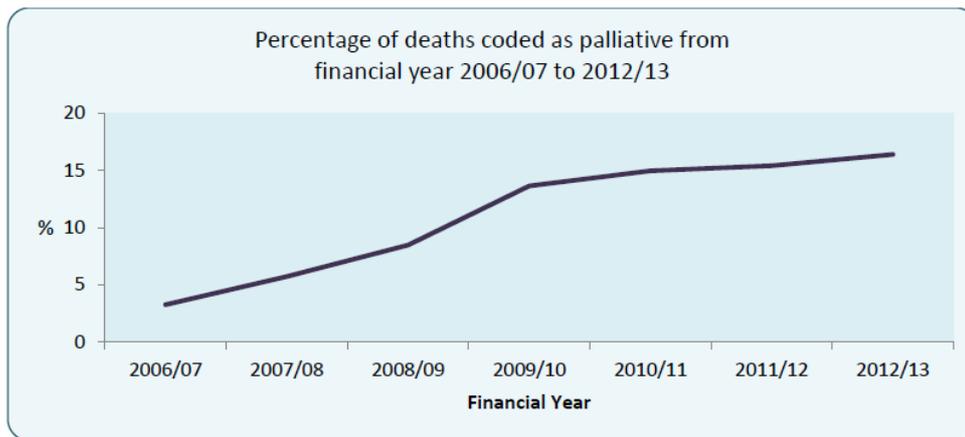
³² CEO Diary, 10 November 2014, Mark Newbold

³³ 'Waiting list accuracy: Assessing the accuracy of waiting list information in NHS hospitals in England', Audit Commission, March 2003

³⁴ 'NHS waiting times for elective care in England', National Audit Office, Jan 2014

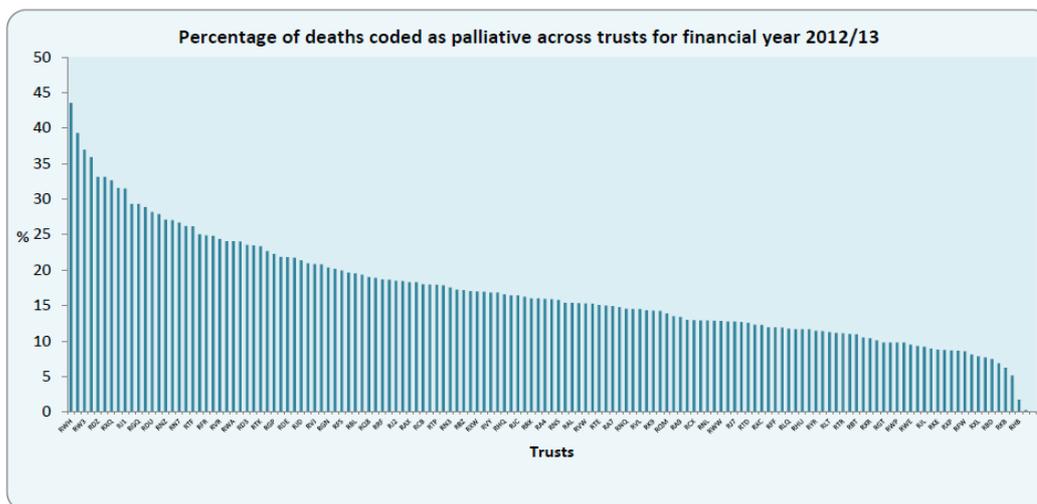
impact on a hospital's reported death rate. Despite this effect being widely discussed, Dr Foster has shown huge variations in the prevalence of palliative care patients and a continued drift upwards over time. Although some of the differences will be due to increasing accuracy of coding and variations in the pattern of local services, such as the availability of home-based and hospice care for terminally ill patients, it is highly likely that there is also an element of gaming the data.

Figure 4 – Percentage of deaths coded as palliative from 2006/07 to 2012/13



Source: Dr Foster

Figure 5 – Percentage of deaths coded as palliative across trusts for 2012/13



Source: Dr Foster

In primary care the QOF was thought by some to lead to gaming. As Gravelle explained, when recording data for QOF:

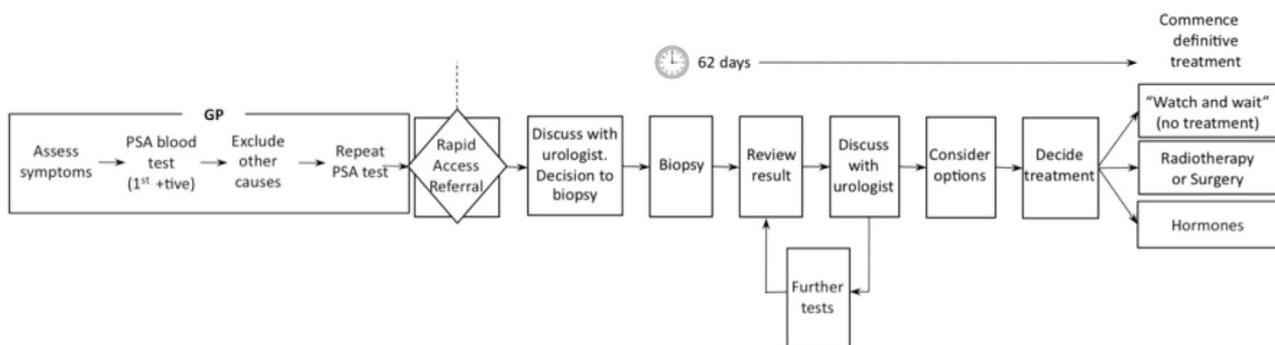
*'Patients can be reported as unsuitable for an indicator on a variety of grounds, for example, if they are terminally ill, frail or cannot tolerate the medication. The practices can also exception-report patients who have failed to attend. So a practice can increase the proportion of patients for whom an indicator is achieved by increasing the number of patients it exception reports.'*³⁵

Distorting the process

The alternative to 'what gets measured gets managed' (tunnel vision) is 'what gets measured gets mis-managed' (distortion). There are numerous reports of changes being made to patient pathways as a result of intense pressure to meet the 4-hour waiting time target in A&E departments. They include: patients being held in ambulances outside hospitals to delay the 'clock starting'; rooms and even corridors being designated as acute observation units so that patients can be categorised as having left A&E; and patients being admitted at the 4-hour point to avoid breaches of the target, where admission could have been avoided if they had been properly assessed.

Other less visible but potentially damaging distortions can be seen in the processes of cancer diagnosis and treatment, where waiting times are also very actively monitored and enforced. In some cases of prostate cancer, where 'watch and wait' (ie no active treatment) and either radical surgery or radiotherapy are all valid treatment options (which if selected will 'stop the clock'), patients may be described as being 'for watch and wait' even when another treatment is actually planned in order to avoid breaching the 62-day treatment target. In some cases, hormone treatment is being initiated earlier than would otherwise be judged most appropriate, for the same reason.

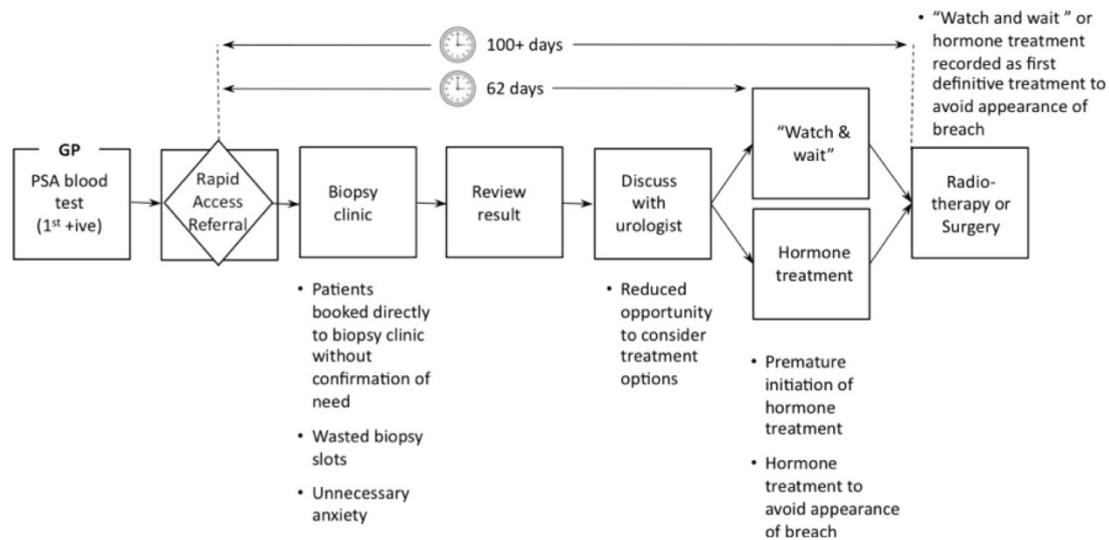
Figure 6 – Preferred patient pathway



Source: Clinician interviews

³⁵ 'Doctor behaviour under a pay for performance contract: Treating, cheating and case finding?' Gravelle et al, *The Economic Journal*, Feb 2010

Figure 7 – ‘Gamed’ patient pathway showing impact of 62-day target



Source: Clinician interviews

6.7. DISTRACTION

Concerns about poor performance often lead to arguments about the data. A range of reactions to data suggestive of poor performance can be observed, such as:

- The data are unreliable
- The sample size is too small
- The type of measure does not apply to us
- It is an isolated example
- It is out of date – things have now improved
- The information does not take into account the very difficult circumstances under which we operate.

In some situations and in certain places at certain times some of these arguments will probably be true. As with many of the adverse effects of performance measurement noted in this section, reviewing and challenging the data is a rational response and can even be a helpful and necessary one, as long as it does not divert attention from investigating the possibility of real performance problems.

Events at Mid Staffordshire again illustrate the dangers of reacting to unfavourable performance information by focusing on arguments about data quality, whilst ignoring underlying service quality and safety issues. In 2007 its

response to data showing high death rates was to establish a group to look into mortality, which, according to the Care Quality Commission, put much of its effort into attempting to establish whether the high rate was a consequence of poor recording of clinical information, rather than finding out why so many patients were dying on its wards.³⁶

'The fact that the organisation concentrated mainly on clinical coding as the explanation for poor outcomes suggests that there was a reluctance to acknowledge, or even consider, that the care of patients was poor.'

6.8. OTHER PUBLIC SERVICES

Health care is not the only sector with issues about performance measurement. Any system in which players are measured based on specific targets carries a risk of gaming. For example, it has been shown that crime statistics have been 'massaged'³⁷ to show police forces in the best light. Targets to reduce crime and/or increase detection rates have been shown to act as incentives for some crimes to be downgraded from notifiable into non-notifiable categories such as anti-social behaviour or crime-related incidents (which are not captured in data returned to the Home Office). As a result, the recorded data does not represent a full, accurate account of crime in England and Wales, leading to an exaggeration of the rate of decrease in crime in some areas.

Education league tables measure schools on how many pupils score from A* to C grade in certain subjects, with maths and English being especially important. A number of changes to the rules around GCSE performance have been made in response to allegations that some schools game the system to improve their rankings in the league tables by:

- Over-grading coursework
- Marking draft coursework and allowing pupils to correct mistakes before submission
- Repeated re-takes to raise grades
- Preventing pupils from sitting GCSEs in subjects where they are weak
- Early exam entry to 'bank' C grades and allow more time for other subjects.

The impact of these practices can be damaging for students, leading to:

- Unrealistically positive picture of their ability, distorting subject choices and leading to longer-term disappointment and demotivation
- Impoverishment of teaching and learning due to cramming for exams

³⁶ 'Investigation into Mid Staffordshire NHS Foundation Trust', Healthcare Commission, March 2009,

³⁷ 'Caught red-handed: Why we can't count on police recorded crime statistics', House of Commons, April 2014

- Lower grades than real ability due to early exam entry
- Discontinuation of learning in a core subject (maths) at age 15
- Lost opportunities for further study

In earlier sections we argued for the necessity and value of measuring and publishing performance in healthcare. The examples in this section have shown how real performance can be obscured, processes distorted, staff ill-treated and service users' interests damaged by the way in which we currently use performance measurement in public services. In the following section we discuss the implications of these types of data abuse, and explore what steps can be taken to minimise them, without abandoning meaningful accountability.

7. Discussion

Healthcare delivery is far more measured, monitored and targeted than ever it used to be. As the previous section illustrates, this has driven important improvements in the standards of care for patients. However, there are examples where the use of performance measures has had detrimental effects for patients, staff and taxpayers.

There is a difficult tension between holding people to account for measured performance and the conditions which encourage open and honest performance reporting. The more serious the consequences of underperformance for individuals, the greater will be the pressure to conceal problems and manipulate data to pretend that things are going well. On the one hand, accountability encourages gaming and can lead to bullying. But measurement without accountability enables poor practice to persist indefinitely, as it did in Bristol, where data showing unusually high death and disability rates in children's heart surgery was not acted on for years.

Meaningful accountability can break down for a whole variety of reasons: if people are set targets that are affected by factors they cannot influence; if they have insufficient resources to do the job; if the scale of consequences for individuals is disproportionate; and if there is only punishment without support for improvement.

In the group discussion about use of data in performance management these issues were all raised. Underlying these comments was a sense that people were being put in unfair or unethical situations as a result of the use of performance data, for example:

- Where it is known that one organisation is getting away with gaming on a performance indicator and not being held to account while another is being penalised for more honest behaviour
- Where the view of a clinician as to the best course of action for a patient conflicts with the requirements of a target – for example, where a doctor is required to prioritise a less urgent patient over a more urgent patient to meet a waiting time requirement
- Where system resources are being diverted towards meeting performance targets without regard to local priorities. At worst, this could mean prioritising achievement of a national target over addressing concerns about sub-standard care
- Where system resources are applied to achieving the minimum necessary to meet performance requirements suggesting disingenuity on the part of those

calling for targets to be met, in that performance over and above the target is accorded no value

- Where little or nothing is done to address known weaknesses in the design of indicators or the format of data used in performance management, particularly when it is known improvements would significantly address some of the unfairness inherent in performance indicators
- Where the individual or organisation is being held to account for a performance measure where they lack the necessary powers to bring about an improvement – for example, where a shortage of social care results in bed-blocking with consequent breaches of 4-hour waiting time targets

These are familiar issues but were raised by many participants in the round table. There was much discussion of the need to make performance management more outcome-focussed, clinically credible, based on local need and meaningful to patients.

These issues are widely recognised and have been behind a number of policy initiatives over the last decade. Some examples are set out in the table below. Efforts in this direction are welcome but they have been met with significant difficulties in implementation, which are also described below.

Participants in the group discussion remained hopeful that further work to improve and develop policy in these areas might yet deliver effective, locally owned, clinically relevant, patient centred performance measurement based on the impact of the whole health and social care system.

Policy area	Examples of relevant policies	Policy challenges
<p>Focus on outcome rather than process measures</p> <p>The best measures in healthcare are outcomes-based. Process measures are inherently dangerous: more open to ambiguous definition, easier to manipulate and more vulnerable to gaming.</p>	<p><i>National Outcomes Framework</i></p> <p>Introduced after 2010, the NOF was designed to focus central government on long-term outcomes and leave the NHS with greater freedom to determine how best to achieve this. It was linked to policy commitments to 'scrap targets'.</p>	<p>As changes in many health outcomes will only be seen over periods that are longer than the political cycle, some dependence on process measures as a surrogate for outcomes is probably unavoidable. The attempt to scrap targets and focus on long-term outcomes was quickly undermined when it became apparent that waiting time performance might suffer as a result.</p>

<p>Use more patient-centred measurement systems Measures that directly reflect patient experience have greater legitimacy. Data that is directly collected from patients avoids some of the potential risks of manipulation.</p>	<p><i>PROMs, National Patient Surveys, NHS Choices, Friends and Family Test (FFT)</i> These initiatives have all created data about patients' direct experience of care.</p>	<p>Robust data is usually high-cost, infrequent and untimely (PROMs, Net Promoter Score). Data that is more frequent and timely is less robust (NHS Choices feedback, FFT). Technologies to deliver both are available but underused.</p>
<p>Clinician-led performance measurement Gaming is encouraged by the perception that a target is clinically irrelevant and indeed may be doing harm. Investment in creating measures that have value to patients and buy-in from clinicians is extremely important.</p>	<p><i>National Clinical Audit</i> Funding for a wide range of clinical audits has been a means to ensure greater clinical buy-in and support for the process of quality measurement.</p>	<p>Participation rates have been variable, as has the quality of data. The data has not been integrated with financial accountability systems, limiting its effective use in performance management. Lack of transparency and control of data by professional groups undermines legitimacy.</p>
<p>Local ownership of performance management The sense that management priorities are dictated from the centre and applied inappropriately and without regard to local circumstance can undermine the validity of measures in the eyes of managers and staff, increasing incentives to game the system.</p>	<p><i>World Class Commissioning and Quality Accounts</i> This aimed to create effective local commissioning organisations able to set local priorities for performance, measure appropriately and hold providers to account. Quality Accounts attempted to give providers room to set their own performance ambitions and monitor their progress against their own goals.</p>	<p>Information and analytical capabilities at the level of local health economies were too weak to deliver WCC at the standards hoped for. Both policies suffer from the conflict of interest that drove some local health economies and providers to create soft, if not positively limp, rods for their own backs.</p>

Measure the whole system not just parts of it

The drivers of individual performance metrics extend beyond organisational boundaries in an institutional structure where competing providers, and commissioners – themselves providers of care – operate within a ‘national’ health service, alongside social care and public health services delivered by locally elected governments. The leaders of local health organisations are thus held accountable for performance measures and targets which are heavily influenced by system effects beyond their control.

Pooling budgets and linking data

Budgets have been pooled through the creation of care trusts – as has happened in Torbay. More is planned: for example the proposed integration of health and social care budgets in Manchester. Care.data will support the integration of data across the health system to allow system-wide performance measurement.

Although mechanisms have been in place, pooled budgets have been adopted relatively slowly. The key challenge to the information aspect of this has been the need to demonstrate privacy protection in the use of linked data sets. In addition there are significant challenges in terms of IT and analytical capabilities to deliver whole system measurement.

8. Recommendations

The roundtable discussion of policy issues that affected the use and abuse of data covered a very wide area, from the organisational design of the health system and financial responsibilities, through to the psychology of individual appraisal and performance management.

The focus of our specific recommendations is the detailed issues concerning the creation and use of performance indicators within that broader canvas. The use of quantitative performance measurement as a tool to improve healthcare is an evolving science in which exploration, innovation, and trial and error continue to yield useful learning. Our survey of experience to date and our discussions with practitioners have led us to five key steps that, if implemented now, we believe would go a long way to reducing abuse of performance data and significantly improving the gain that could be achieved through its use.

8.1. MAKE DATA QUALITY AS IMPORTANT AS HITTING TARGETS

The whole enterprise of data-led performance management and transparency is perpetually challenged by inaccurate data. Although multiple audit reports over a period of years have highlighted problems with data quality, long-term solutions remain elusive.

The failure of the NHS to adequately address this issue calls into question its commitment to fair and effective performance management. This can be seen most clearly in the imbalance between the level of seriousness with which poor performance on targets is treated, compared to the effort expended on identifying gaming and inaccurate recording of information.

There exists a conflict of interest within the performance management regime whereby exposing the unreliability of data showing improved performance will often create as many problems for the performance manager as it will for the organisation being managed.

The foundation for creating high-quality data is ownership by the people who use it, especially clinicians and patients. The data needs to be actively used, analysed, audited and curated, and organisations need to invest the relatively modest resources to do this. In practice, the unglamorous back-office work of managing, coding, checking and cleaning information tends to attract attention only when miscoding can be blamed for apparently poor performance, whilst the quality of data underpinning average or strong performance tends not to be examined.

Wherever possible, performance measures should be created using information whose production is beyond the control of the organisation being monitored. For

example, patient reported outcomes and survey data are more reliable because of their independence. Similarly, clinical data should, ideally, be drawn from sources that are shared with patients and used in other contexts. The recent announcement about sample case note reviews to quantify avoidable death rates is an example of performance reporting where inadequate thought has gone into the assurance of the quality of the information that will be produced. Such a major step to extend an academic technique into practice in every hospital is bound to create implementation challenges and may well have unintended side effects.

Overall, current performance management and inspection regimes make disproportionately minimal efforts to audit and assure the quality of information that underpins the entire system. Data assurance audit in the NHS has waxed and waned with changes in perceived priorities and restructuring of the public audit environment. Now that the National Audit Office has taken over responsibility for overseeing the audit of local services including the NHS, there is an opportunity to revisit the emphasis and resources devoted to auditing data quality. It would be extremely valuable for the National Audit Office to commit to an ongoing audit programme on data integrity, recognising that constant attention to data quality will continue to be needed over the long term.

8.2. MEASURE THE CONTEXT NOT JUST THE INDICATOR

Performance management frameworks need to comprise sets of counterbalancing metrics for every target. These counterbalancing measures should be designed to guard against negative consequences for patients. In the case of 4-hour waits in A&E, for example, the appropriate measures would include ambulance queueing times; re-attendance rates at A&E; the average risk scores of admitted patients or the number of patients with bottom quartile risk scores; and numbers of short stay admissions. A strong performance on A&E 4-hour waits is of limited value if accompanied by high numbers of short stay admission for low-risk patients or high levels of discharge and re-attendance.

In complex and dynamic systems like the NHS, counterbalancing metrics may not be obvious initially, and may need to change as the system adapts to optimise around new targets. This implies that measures themselves need to be constantly monitored and reviewed in the light of experience. The link between selected process measures and outcomes needs to be established. Performance metrics could be designed, owned, benchmarked internationally and actively curated by multi-disciplinary specialist groups, including royal colleges and patient organisations, taking responsibility for observing implementation and identifying and investigating unintended effects, especially in the early years. These groups could take a lead in using the data to catalyse collaboration and competition between professionals and institutions, to aid performance improvement.

Lastly, performance management regimes should recognise the risks created by measuring some aspects of the system, to the quality of areas not measured. In particular, measurement 'black holes', such as community services, must be comprehensively identified and actively addressed.

8.3. AVOID THRESHOLDS

In the same way that process measures are more open to manipulation and gaming than outcomes, the imposition of threshold measures to continuous variables tends to cause problems. In the case of waiting times, for example, shorter is generally better. But despite the fact that there is rarely a magic number with evidence of a significant difference in patient outcomes on either side of the line, the existence of the target tends to influence the way resources are deployed and to skew decision-making. It would mitigate some of the negative consequences of performance management to present results as a continuous variable rather than as pass-fail against an artificial threshold. This would make performance differences visible without the distorting effects of the threshold. The data could still be used in a variety of different ways by practitioners, patients and commentators, and could help inform management judgements and direct external intervention to where it is most needed.

Thresholds are appealing because they make the implementation and communication of performance management simple. However this needs to be balanced against risks of negative consequences. This might be done, using waiting times as an example, by designing a regime that rewarded reduction in average waiting times while imposing thresholds for the maximum acceptable waiting time for any individual patient. There is a particular problem with outcome measures where statistical approaches favour categorising performance as either 'outlying' or 'not outlying'. While this is valid statistically, and useful in some contexts, it makes little sense in a performance management regime to treat performance one side of a statistical threshold as no cause for concern and performance that is not significantly different but which lies the other side of a threshold, as being a major problem.

8.4. BE OPEN

The data used for performance management should be made available at the lowest possible level of granularity. It should be provided to researchers and independent analysts to allow indicators to be tested and issues to be identified. Transparency of data is an essential aid to system learning and development. To give one example, in preparing this paper we wished to investigate the degree to which QOF targets had created ceilings in terms of performance – where the existence of a target reduces incentives to improve beyond the targeted level. However, the data necessary to test this proposition is not in the public domain. Similarly, the use of clinical audit data that is not subject to external scrutiny undermines the validity of performance management based on such information.

An ongoing and active programme should be established to encourage research into the impact of performance targets, identify unintended consequences and propose continuous improvement in the metrics. Data created through the NHS would be an enormously valuable resource in this enterprise and would help keep the UK at the forefront of health system performance measurement.

8.5. APPLY MEASURES FAIRLY

An approach to performance management that regards any failure to meet targets as a problem is inherently unfair because in some instances, the right thing to do for patients and for the NHS is to prioritise something over and above targets. If the performance management regime is incapable of recognising when such a situation has occurred, it will lack legitimacy.

The difficulty for performance management is to allow such a dialogue to occur without being drowned by a cacophony of objections and special pleading. However, to shut down all and any such conversation is equally problematic.

The same considerations apply to the publication of information. Apparently good performance may be the result of inaccurate data, deliberate falsification, or distortion of care processes against the interests of patients. There has been much focus on the need to temper the language of media in terms of description of poor performance and to acknowledge that appearances can be deceptive.

In contrast, there has been insufficient work within the performance management system to distinguish between genuinely poor performance correctly identified by targets, as opposed to performance mistakenly identified as poor because of weaknesses in the information. At the very least, there must be the possibility of such a judgement being made if the regime is to be considered fair.

By the same token, failure to examine the performance of organisations that appear to be doing well carries significant hidden costs. Where success is genuine, valuable learning about innovative and successful approaches may be overlooked. Where the data does not fairly reflect the real standards of care, patients may be suffering. There is a strong argument for managers and regulators to examine in depth, and with a critical eye, apparently strong as well as weak performance.

9. Acknowledgement and thanks

This report was produced with the help of the following people whose contributions helped inform the debate. However the views expressed are those of the authors alone.

1. The Dr Foster Ethics Committee

- Joanne Shaw, Chair – Chairman of Datapharm and Deputy Chair of Nuffield Health
- Dr Hugo Ford – Director of Cancer Services, Cambridge University Hospitals NHS Foundation Trust
- Mr Phil Hope – Adjunct Professor, Global Health Institute, Imperial College, Co-Director of Improving Care, and former Labour Minister of State for Care Services
- Dr James Kingsland OBE – Senior Partner in a Personal Medical Services Partnership in Wallasey, National Clinical Commissioning Network Lead for England
- Mr Ian Lane MA, DM, MCh, FRCS – Former Medical Director, Cardiff and Vale NHS Trust
- Professor Barry McCormick CBE, PhD – Director, Centre for Health Service Economics & Organisation, Nuffield Department of Primary Care Health Sciences University of Oxford
- Mr Brian Ridsdale – Actuary, UK Representative, International Actuarial Mortality Working Group
- Dame Jane Roberts – Former Director for Quality and Performance, Islington Primary Care Trust
- Dr Patricia Wilkie, OBE – President and Chairman of the National Association for Patient Participation

2. The expert reference group

- Nick Carver – Chief Executive, East and North Hertfordshire NHS Trust
- Dr Rupert Dunbar-Rees – Founder & CEO, Outcomes Based Healthcare
- Dr Murray Ellender – GP Partner, Medical Director Urgent Care, The Hurley Group
- Dr Benjamin Ellis – Senior Clinical Policy Adviser, Arthritis Research UK

- Dr Martin Farrier – Consultant Paediatrician, Associate Medical Director, Wrightington Wigan and Leigh NHS Foundation Trust
- Dr Hugo Ford – Director of Cancer Services, Cambridge University Hospitals NHS Foundation Trust
- Dr Thomas Kelley – Europe Director, International Consortium for Health Outcomes Measurement
- Maggie Kemmner – Programme Director, Southwark and Lambeth Integrated Care
- Associate Professor Daniel Lasserson – Nuffield Department of Primary Care Health Sciences, University of Oxford and Oxford Academic Health Science Network
- Jacque Mallender – Partner, Public Policy, Optimity Matrix
- Marcus Powell, Group Organisational Development and HR Director - Nuffield Health
- Dr David Rogers – Medical Director, Cumbria Clinical Commissioning Group

3. Dr Foster team

- Rachel Alsop
- Nisha Rajendran
- Kayleigh Pearse

Copyright © April 2015. Unless explicitly stated otherwise, all rights including copyright in the content of this report are owned by Dr Foster Limited. The content of this report may not be copied, reproduced, republished, posted, broadcast or transmitted in any way without first obtaining Dr Foster Limited's written permission. This report may be referenced or quoted (but not for commercial purposes) as long as copyright is acknowledged.