Why Risk-based Regulation of Healthcare Quality in the NHS Cannot Succeed

Anne-Laure Beaussier,* Department of Geography, King’s College London, Strand, London WC2R

David Demeritt, Department of Geography, King’s College London, Strand, London WC2R 2LS

Alex Griffiths, Care Quality Commission and Department of Management, King’s College London

Henry Rothstein, Department of Geography, King’s College London, Strand, London WC2R 2LS

*Corresponding author: anne-laure.beaussier@kcl.ac.uk

HowSAFE working paper 5

5 August 2015

Research for this Working Paper was supported by a grant awarded through the Open Research Area Programme for the Social Sciences jointly funded by the ESRC (ES/K006169/1), DFG, ANR, and NOW (http://tinyurl.com/howsafe-project).
Why Risk-based Regulation of Healthcare Quality in the NHS Cannot Succeed

Abstract. This paper explores the challenges of regulating healthcare quality and explains why risk-based approaches have not and cannot solve the problems of governing the National Health Service (NHS) in the UK. Based on historical policy analysis and in-depth interviews with 15 high-level informants, it documents the instability and layering of ambiguous and often contested and contradictory definitions of healthcare quality in the NHS. Partly as a result regulators have also struggled to measure quality and to identify which providers are failing to meet the required quality standards, despite world-leading systems for statistically analysing the wealth of outcomes data and indicators generated by the NHS. There are also problems with applying the conventional enforcement pyramid of ‘carrots and sticks’ in healthcare because deterrence-oriented sanctions in response to non-compliance risk further undermining care quality and more compliance-oriented levers for enforcing quality standards have proven ineffective in a sector characterised by strong information asymmetries and monopolistic provision. While risk-based approaches promise to help rationalise and manage the inevitable limits of governance and what it can hope to achieve, we argue that they will inevitably disappoint so long as there is no political tolerance for failure in the NHS. Beyond these difficulties in defining acceptable quality standards and assessing how far it is reasonable to go in trying to meet them, we also raise questions about the extent to which ensuring healthcare quality in the NHS is a problem that even admits of regulatory solutions at all, given the systemic problems of delivering a service in the context of numerous conflicting governance goals and constraints.

Key words: risk-based regulation; healthcare governance; quality standards; Care Quality Commission; National Health Service (NHS)

1. Introduction

Over the last thirty years the desire to improve the quality of healthcare has been the one constant amidst near ceaseless reform to the British National Health Service (NHS). Whereas Thatcher-era reforms looked to dedicated managers and to the discipline of the internal market “to secure the best deal for patients and the community health within available resources” (R. Griffiths, 1983, para. 3), subsequent governments created independent watchdogs, charging a succession of steadily more powerful regulatory agencies with monitoring quality and ensuring standards. But the experience of these agencies has not been a happy one. Stronger agencies and sophisticated systems for risk assessing and managing the quality of healthcare have neither calmed chronic public concern nor quelled political controversy over NHS waiting lists, patient safety and care standards.

While some of these challenges are unique to healthcare, NHS managers are far from alone in struggling to meet demands for more cost effective, accountable, and responsive public services. From teaching and social work to banking and the bar, once autonomous and entirely self-regulating professions are increasingly subject to external oversight from government (Moran, 2003). These changes in the organisation of public services reflect a more general shift in the role of the state from ‘rowing’ and direct service provision, to ‘steering’ and regulation (Osborne & Gaebler, 1992). Thus while the NHS in England remains publicly funded and free at the point of delivery, service provision is no longer under ministerial control. With NHS Trusts becoming autonomous Foundation Trusts and care increasingly provided by the private sector, the government has come to
Why Risk-based Regulation Cannot Succeed

rely more on external inspections by independent regulators to ensure basic quality and standards. In common with many other public services, however, external regulatory inspection and oversight has proved both costly and controversial, with critics highlighting the perverse consequences for care quality of ham-fisted regulation (Barker, Pearce, & Irving, 2004; Bevan & Hood, 2006).

In response to such difficulties, regulators have- over the last decade or so- adopted various risk-based approaches to organizing regulation in order to improve its effectiveness and proportionality (Rothstein, 2006). Rather than trying to prevent all possible harms, ‘risk-based’ approaches to regulation focus regulatory standard-setting, information gathering, and enforcement activities on controlling the highest priority harms, as determined through ex ante assessments of their probability and consequences. Such approaches have been enthusiastically promoted and, indeed mandated, by successive UK governments as a universal principle of ‘better regulation’ (Demeritt, Rothstein, Beaussier, & Howard, 2015).

In the particular case of healthcare, risk-based approaches have at least two further appeals (Phipps, Noyce, Walshe, Parker, & Ashcroft, 2011). First, medical interventions can often be dangerous to patients, and so ensuring care quality, at both the individual and population levels, requires careful consideration of the risk-benefit trade-offs that are central to risk-based regulation. Second, unlike the risk-based approaches developed for regulating quality in other sectors such as higher education and social care, NHS regulators can use the comprehensive outcomes data generated by the NHS’s single payer system to underpin their risk-based regulatory model. Thus, over the last decade, the current healthcare regulator- the Care Quality Commission (CQC)- and its predecessors have sought to target their regulatory interventions by investing heavily in a remarkable world-leading system for statistically measuring care quality and identifying those providers at greatest risk of failing to meet the required standards.

For all its technical sophistication and considerable cost- totalling approximately £16m/year for the analytical staff alone (CQC, 2014b, p. 31)-, this risk-based approach has been unable to prevent significant failures in care quality, or to insulate the CQC from the consequent scandals. Revelations of systemic neglect and negligence at Stafford Hospital, the abuse of learning-disabled adults in care at Winterbourne View, and the death of 11 babies and a mother at Morecombe Bay NHS Foundation Trust sit uneasily alongside the cherished image of the NHS in the national consciousness. Indeed, accusations of incompetence, failure and cover-up by the CQC led to the resignation of two CQC chairs and a chief-executive in 2009 and 2012. After a further round of regulatory reform, the NHS is in trouble again, with the CQC rating more than 80% of acute hospital trusts inspected in England over the last two years as ‘inadequate’ or ‘requires improvement’ as it struggles to cope with over 90,000 safeguarding alerts a year (2015a).

In this paper we explore the challenges faced by risk-based regulation of healthcare quality and explain why it struggles to succeed in the NHS. The paper follows a qualitative methodology combining policy document analysis and in-depth interviews with 15 high-level informants closely involved in the design, implementation, and reform of successive care quality regulators. After briefly discussing the history of care quality regulation, we will discuss the internal and external challenges facing the three functional components of the care quality regime- that of defining the goals of the quality regime and assessing and enforcing compliance with those quality goals- which, from a control systems perspective- any fully functioning regulatory regime should possess (Hood, Rothstein, & Baldwin, 2001).
2. History of care quality regulation in England

Since the founding of the NHS in 1948, institutional arrangements for governing the quality of healthcare can be roughly classified into four distinct phases, each introducing new institutions, logics, and styles of governance. The first phase was based on professional self-regulation. While the NHS nationalised healthcare delivery, medical professionals jealously guarded their authority over the quality of medical practice and general care, which as Klein has observed, reduced NHS governance to little more than a system of “exhort and hope” (Klein, 2010, p. 285). High-level oversight, standards and guidance for medical practice and training were provided by the centuries-old royal medical colleges, and a simple statutory register of medical practitioners was kept by the General Medical Council. Day-to-day oversight of care quality, meanwhile, was left to doctors and clinical teams who were responsible for reviewing their own performance, albeit subject to a complex complaints procedure. As Klein notes, there was no “systematic performance evaluation ... [nor did] it feed into the mainstream of bureaucratic evaluation, since any information generated about performance... [was] restricted to the medical domain” (Klein, 1982, p. 405).

A second phase of new public management-inspired approaches to governing healthcare quality began in the early 1980s. While professional standards remained the dominant yardstick of clinical quality and best practice, the scope for professional autonomy was increasingly restricted by pressures on the NHS to demonstrate value for money. In 1983, the Conservatives introduced a national ‘performance’ indicator regime, which drew on existing management data to measure 70 indicators of inputs, care volumes, and cost effectiveness to inform service planning and provision. Over the next twenty years, the Department of Health (DoH) introduced some 3000 further indicators to target much wider issues of access, clinical outcomes and, more recently, patient satisfaction. These indicators were made public in the late 1980s and then published as league tables in 1994 (Pollitt, Harrison, Dowswell, Jerak-Zuiderent, & Bal, 2010), reflecting increasingly acknowledged patient rights, illustrated for example by the 1992 Patients’ Charter and the 2009 NHS Constitution.

With the introduction of the internal market under the 1990 National Health Service and Community Care Act, indicators assumed particular significance as a means of ‘steering’ rather than ‘rowing’ the healthcare system. That Act sought to drive improvements in the efficiency and quality of healthcare by introducing competition among NHS providers, purchaser-provider contracting, and later on a ‘money follows the patient’ funding regime. While the Act freed central government from day-to-day management of the NHS, the DoH used the indicator regime to inform the new contracting arrangements and hold the newly created NHS trusts to account for their financial management and healthcare delivery (Pollitt et al., 2010). That regime, however, did not explicitly address the standards of care individual patients received, which was left to expert clinician judgment and professional self-regulation.

The third phase of care quality regulation was marked by the introduction of ‘clinical governance’ under the 1999 Health Act, when the New Labour administration imposed a statutory Duty of Quality on NHS Trust Boards. That new duty emerged in response to rising public expectations about service standards, increasing complaints and litigation against NHS trusts, and some high-profile scandals including the deaths of several children from heart surgery at the Bristol Royal Infirmary, the illegal retention of childrens’ organs at Alder Hay hospital, and the murder of 200 patients by the GP Harold Shipman (Alaszewski, 2002; Bevan, 2008). Borrowing from fashionable ideas of corporate
Why Risk-based Regulation Cannot Succeed

governance, clinical governance sought to change the culture of healthcare by holding professional judgement to account through quality assurance processes.

Those processes largely entailed checks against national standards set by professional bodies and the increasingly influential National Institute for Clinical Excellence (NICE), whose risk-benefit analyses shaped the DoH’s own National Service Frameworks in determining what treatments would be available on the NHS. Performance was monitored through the national NHS Performance Assessment Framework, which introduced a ‘balanced scorecard’ approach to weighing a set of key metrics against DoH objectives for health improvement, efficiency, accessibility, and patient experience (Chang, Lin, & Northcott, 2002). The 1999 Health Act also created the Commission for Health Improvement (CHI), as a non-departmental public body to review how NHS trusts implemented clinical governance. Although CHI had no autonomous sanctioning powers, it did organise rolling peer-reviews of all NHS trusts. These visits involved ad hoc multi-disciplinary teams led by a CHI review manager, including a nurse, doctor, NHS manager, lay member and another clinical professional all recruited through national advertisement and trained for the visits (Bevan & Cornwell, 2006, p. 10). CHI reviews identified serious failures in several hospitals, ranging from 20-hour trolley waits and filthy toilets to high death rates, prompting a number of chief executives to resign (Nuffield Trust, 2013). Inspired by Ofsted’s muscular approach to naming and shaming bad schools, the DoH decided, in 2001, to complement CHI clinical governance reviews by publishing annual ratings to inform the new patient choice agenda. Acute hospital trusts were awarded zero to three stars based upon performance against key national targets, such as waiting times. CHI took over the star rating system the following year and by 2003 had applied it to all types of NHS trust.

The 2003 Health and Social Care Act replaced CHI with a more powerful inspectorate, the Healthcare Commission (HCC), as part of a wider reform aimed at increasing capacity through competition and involving private healthcare providers. The HCC was given enhanced enforcement powers to assess healthcare performance across its expanded remit of private and NHS healthcare providers. To that end, the HCC replaced ad hoc teams of peer reviewers with a dedicated professional inspectorate and developed an innovative monitoring system using administrative data to detect statistical abnormalities in hospital performance across a wide range of indicators and help target inspections in proportion to risk of non-compliance (Bevan & Cornwell, 2006). After two years, the HCC discontinued the star-ratings and clinical governance reviews it had inherited from CHI in favour of a new ‘annual health-check’ approach to regulatory inspection. This approach assessed Trusts against a set of core national standards, publicly awarding one of four ratings from ‘weak’ to ‘excellent’ based on the Trust’s own self-assessment document, which the HCC verified using its own statistical monitoring system and on-site inspections (Nuffield Trust, 2013).

The fourth, more fully risk-based phase of healthcare quality regulation emerged from the 2008 Health and Social Care Act, which replaced the HCC with the Care Quality Commission (CQC), bringing together quality regulators from across the health, social and mental healthcare sectors. The consequences for the merged regulator were dramatic. Whereas the first regulator, CHI, had been responsible for monitoring care quality in 200 acute and specialist NHS trusts, the CQC was responsible for registering, and assuring care quality outcomes in 44,000 organisations providing primary, acute, mental health, and social care across England.

To prioritize those increased duties, the CQC built on HCC’s statistical approach to monitoring NHS hospitals by developing what it called ‘quality risk profiles’ (QRPs), which were essentially...
assessments of the risk of regulatory non-compliance, for all health and social care organisations. As inspection resources became scarcer, the CQC made more extensive use of quantitative data to develop its risk-assessments and relied far less on on-site checks, which decreased sharply between 2009 and 2011 (NAO, 2011). The new approach was based on analysing 1000 outcome measures-comprising 1.2m individual quantitative, and 100,000 qualitative, data items (A. Griffiths, 2012)- that were then matched against 16 essential standards established by the DoH. The CQC also merged the various dedicated inspectorates it had inherited into a single generalist inspectorate, whose staff were no longer required to have any expertise in the various fields they were inspecting. CQC inspectors were supposed to use the QRPs to “prioritise which services we inspect, and when, how, and what we will focus on in the inspection” (CQC, 2014a). Within nine months, however, its first Chair resigned after the CQC rated Basildon hospital as ‘good’, only to contradict itself weeks later by taking enforcement action following a critical inspection report. Then, in 2012, the second Chair of the CQC and its Chief Executive both resigned following damning revelations about its risk-based inspection regime from CQC whistleblowers.

Yet rather than being replaced or even abolished by a government otherwise committed to a ‘bonfire of the quangos’, under the 2012 Health and Social Care Act the CQC gained more autonomy from the DoH, enhanced enforcement powers and an increased budget. The Francis Inquiry (2013) into the failings at Mid-Staffordshire NHS trust led to further reforms of the CQC’s essential quality standards, a simplification of its statistical methods of ‘Intelligent Monitoring’ to assess the risk of non-compliance (CQC, 2014d), and a complete overhaul of its system of inspection and enforcement (CQC, 2015d). The current system now comprises three specialist inspectorates for hospitals, social care and general practice led by high-profile and experienced ‘Chief Inspectors’. The CQC is also undertaking a transitional programme of systematic inspection of all acute hospitals pending the development of yet another risk-based model for prioritising future inspections. While the revised regulatory model has received a generally favourable early response (House of Commons, 2014), particularly from acute hospitals (Walshe, Addicott, Boyd, Roberston, & Ross, 2014), there has already been some disquiet from GPs (Millett, 2015).

In the sections that follow we explore the challenges to adopting a risk as a central organizing principle for defining, assessing, and enforcing care quality regulation in the NHS.

3. Defining Quality

One reason why risk-based regulation of the NHS has struggled to live up to its promises is that definitions of care quality, and thus the goal of regulation, have been unstable. Although some medical experts insist “quality can be defined and measured” objectively (Chassin MR, Galvin RW, & the National Roundtable on Health Care Quality, 1998, p. 1001; cf. Mainz, 2003), in practice definitions of quality have changed significantly over time and rarely attracted consensus amongst the policy and medical communities or wider publics. Since the creation of CHI fifteen years ago, the DoH has substantially redefined care quality frameworks and standards at least five times.

This constant change is perhaps not surprising. As Donabedian (1988, p. 1743) has argued, it was not that long ago that the definition of quality was “considered to be something of a mystery: real, capable of being perceived and appreciated, but not subject to measurement.” Over the last few decades, campaigns to improve care quality in the NHS have variously focused on regulating different aspects of Donabedian’s influential healthcare ‘triad’ of healthcare ‘structures’ (e.g.
facilities, staffing, budgets, and capacities), ‘processes’ (e.g. care procedures, systems and organisations), and ‘outcomes’ (e.g. clinical effectiveness, patient and public health) as proxies for quality itself. But far from entailing a teleological evolution towards a more complete definition of quality, these different approaches have been layered on top of each other, reflecting the changing priorities of different governments and even health secretaries in balancing the often competing goals of treatment volume, cost containment, clinical outcomes, and patient satisfaction.

Thus one marker of ‘good’ healthcare throughout the history of the NHS has been the quality of healthcare infrastructures; the white coats, buildings and budgets necessary to deliver ‘free at the point of delivery’ healthcare. New public management efficiency drives, from the 1980s onwards, embodied by the introduction of volume-based targets, did not fundamentally alter this approach insofar as they were focussed on structural aspects of quality.

Clinical governance, by contrast, offers a rather different approach to quality, based on sound processes and coherent organisational systems. This process-based approach to quality regulation was layered on top of structural efficiency drives towards the end of the 1990s, at least partly in response to concerns that volume-based targets were compromising patient safety. These two approaches coexisted alongside each other in the early 2000s, with clinical governance audits operating in parallel to a national target regime primarily aimed at reducing waiting lists. In a sense, the star-ratings scheme of the early 2000s was an attempt to reconcile the goals of volume and safety, but fundamental tensions remained; notably illustrated by high-profile struggles to contain hospital acquired infections such as MRSA while meeting high bed occupancy targets (Cunningham, Kernohan, & Rush, 2006).

Measurable outcome-based approaches to quality were gradually implemented in the following decade, adding yet another layer to the quality ‘cake’. These approaches to quality have been defined through indicators of patient safety and clinical effectiveness such as standardised mortality or readmission rates. Increasingly, however, outcome indicators have also related to subjective patient experience- such as the ‘friends and family’ test- reflecting the classical business management nostrum, ‘quality is what the customer says it is’. As the current President of the BMA put it to us, “We’ve had lots of written standards and criteria and ... everybody drowns under guidelines. But actually at the end of the day, the attitude is- is it good enough for my mum or dad or son or daughter or brother or sister? If it isn't, then it's not good enough for anybody else either”.

That focus on patient experience has been reinforced through the idea of ‘compassionate care’, developed in response to criticisms about the unresponsiveness of the NHS, and further advocated by the Francis Report (2013). Now, in addition to professional certification and licencing, the DoH (2015) requires all health and social care workers to complete training and obtain a Care Certificate to demonstrate that they “can meet each of the 15 [compassionate care] standards”, including inter alia “caring with privacy and dignity, awareness of mental health, communication and infection control”. Rather than supplanting previous desiderata of care quality, compassionate care has been simply been layered on top.

While definitions of quality have constantly expanded, not least in response to the latest scandal, regulators have paid less attention to the inevitable trade-offs between them. Thus, the CQC’s current standards- safe, effective, caring, responsive and well-led- could easily be read as a set of conflicting maxims. For example, expectations of ‘well-led’, which entail socially optimal balances
between treatment and costs, sit uneasily with the more maximalist implications of ‘responding to people’s needs’, especially if individual patients and their doctors conceive of treatment as an absolute and unqualified goal. Likewise, ‘effective’ and ‘safe’ hospitals may well be the big ones that can concentrate expertise and benefit from economies of scale, but local communities, politicians and patient associations may well resist the closure of what are regarded as ‘caring’ but less efficient or even less safe local hospitals (Timmins, 2007). Those trade-offs were likely to be made even harder by the Government request in 2015 that the CQC also assess the ‘effective use of resources’ on top of the other five essential quality standards it had so painstakingly redefined just two years before (CQC, 2015f).

With their emphasis on qualified, rather than absolute governance goals, risk tools could in principle help manage those trade-offs insofar as they use _ex ante_ assessments of probability and consequence to help determine how far it is reasonable to go in order to avoid potential adverse outcomes. However, this rationalizing logic has not always resolved health policy debates, which involve contested normative principles. While the idea that medical treatment entails inherent risk may be broadly accepted, in practice regulators and the NHS have struggled to agree _ex ante_ about how to distinguish unacceptable failures from the inevitable but acceptable risks inherent to medicine. The concept of ‘never events’ illustrates this unease. For example, wrong site surgery is never supposed to happen, but it occurs with alarming regularity. In 2009, the first year in which they were systematically recorded, the CQC reported that there were 111 ‘never events’ across the England (CQC, 2014c). Although formally unacceptable, ‘never events’ are in practice tolerated because the measures required to eliminate them, such as abandoning surgical interventions, are even less acceptable.

Even when regulators and the NHS can agree on what is _ex ante_ acceptable, the high public saliency and political sensitivity of the NHS mean that there is often little tolerance for adverse outcomes when they come to light. For example, when hospital acquired infections burst onto the headlines in the mid-2000s, the government set reduction targets, but politicians had little appetite for convincing the public and the media that it would be disproportionately costly to eradicate such infections completely. Despite expert calls to “stop pandering to populism about hospital cleanliness” (The Lancet, 2007) and to focus instead on the more critical factor of bed occupancy rates (Cunningham et al., 2006), the Conservative opposition pledged to make “ridding our hospitals of the superbug a key priority” (Howard, 2004) and the Labour government responded with a £50million pledge to ‘deep clean’ every hospital ward in the country (Watt, 2007). Such political interventions tend to undermine whatever _ex ante_ compromises had been previously made between competing quality goals and to reinforce the tendency for new definitions and standards of care quality to layer-up on top of each other in unstable configurations that are consequently difficult to assess and enforce.

4. **Assessing Quality**

Even with widely accepted and unambiguous definitions of healthcare quality, risk-based regulation faces further challenges in devising appropriate measures for assessing quality and identifying which providers are at greatest risk of failing to meet the required standards. Traditional forms of professional self-regulation rely on the tacit knowledge and subjective expert judgments of fellow professionals to assess quality and compliance with best practice. However, the insider knowledge
that qualifies professionals to judge their peers also tends to compromise the independence and impartiality of the resulting quality assessments. For example, the Bristol Royal Infirmary Public Inquiry cited the hospital’s "club culture" as one reason why abnormally high mortality rates among paediatric heart patients went undetected for so long (Kennedy, 2001, para. 8).

The regime of clinical governance established in response to that scandal introduced quality assurance and external audit requirements designed to make quality assessment more independent and impartial. To that end, the first clinical governance reviews by CHI were conducted by *ad hoc* teams of senior consultant doctors, while successor agencies employed their own full-time inspectors. These forms of external scrutiny were more independent and less prone to the partialities of self-regulation and peer review, but they were still reliant on subjective judgment; making their assessments of quality only as discerning as they were expert. One of the central criticisms levelled against the CQC’s merger of health and social care inspection in 2008, was that its full-time inspectors were generalists who lacked the expertise needed to judge quality across their wide ranging remit (Francis, 2013). Hospitals, for example, encompass a wide range of clinical specialties that would be hard for any healthcare professional to assess skilfully, particularly in the face of resistance from health chiefs anxious to avoid external scrutiny as happened at Mid-Staffordshire Trust (Carvel, 2012). But generalist inspectors found it even harder to inspect a hospital credibly if they had a background in social care or policing, or were one of the 134 inspectors revealed as employed during chronic staff shortages despite failing recruitment tests and “without the core competencies to do the job” (Hazell, 2014)!

To supplement the personal judgment of their inspectorates, healthcare regulators have also developed a series of statistical indicators for monitoring quality, in the firm belief that “we can only be sure to improve what we can actually measure” (DoH, 2008, p. 49). The first NHS Performance Assessment Framework sought to distil a set of high level performance indicators from the wealth of administrative data generated by the NHS. Reflecting their source, these first indicator sets focused on the supply side, measuring costs, care volumes, waiting times, and health outcomes at both patient and population levels, rather than subjective patient experience. In 2001 the government addressed that gap by commissioning an annual survey of hospital inpatients to inform hospital ratings. Since then, the expanding patient opinion industry has spawned ‘Patient Reported Outcome Measures’, the much maligned ‘Friends and Family’ test, and a dedicated consumer champion quango, Healthwatch England (Kmietowicz, 2013).

While statistical methods of quality assessment promise greater objectivity, transparency and therefore trustworthiness than the subjective judgments of professional peer reviewers and external inspectors (Porter, 1995), critics have highlighted “formidable difficulties in organizing the regulation of the quality of healthcare using the data that are routinely available” (Bevan & Cornwell, 2006). The first generation of indicators used by the CHI were criticised for focusing on process measures, such as using prescribing rates or implementation of out-of-hours phone access plans for monitoring Primary Care Trusts (CHI, 2003), rather than necessarily capturing what matters most for the quality of care (Nuffield Trust, 2013). The HCC and, later, the CQC based their assessment on a mixture of process and outcomes measures, in an attempt to more directly assess quality, but developing valid indicators has proved to be technically challenging in at least five ways.

One difficulty relates to the level at which monitoring takes place. Clinical data is typically reported at the Trust-level, rather than at the site-specific level where care is actually delivered. As a
result poor outcomes in one service may be masked by good outcomes elsewhere in the Trust. Timing is a second challenge as data release can lag many months behind, reducing its value for identifying which providers are currently at greatest risk of noncompliance.

A third challenge relates to the metrics for benchmarking performance. Outcomes data, for example, must be normalised to take account of the varying patient mixes using different sites and services. While quantitative techniques like winsorised z-scoring have provided a way to compare performance across different statistical units (Spiegelhalter et al., 2012), their use subtly transforms quality assessment from a judgment of compliance with absolute standards into one of relative performance. For example, a Trust with an alarming 1,000 cases of MRSA one year will appear as the safest in the country if every other Trust has 1,001. ‘Lowest relative risk’ may not be the same as ‘acceptable risk’ when judged against absolute standards of ‘quality care’

A fourth challenge is indicator interpretation. Standardised mortality rates, for example, have long been used to assess hospital performance because they are held by their advocates, such as Dr. Foster, to provide a useful ‘smoke alarm’ for warning of potential quality problems (Donnelly & Sawer, 2013). Critics, however, argue that such rates do not consider how many of those ‘excess deaths’ were preventable and thus “should not be used to benchmark hospitals’ quality of care” (Ramesh, 2015). In 2015, the Secretary of State announced that the NHS would be adding an ‘avoidable death’ indicator to existing outcomes data, but without intensive expert and qualitative assessment it is unclear how this will be achieved.

A fifth technical challenge is the sheer number and diversity of indicators, which means that it is highly likely that all Trusts will be identified as posing a risk on one indicator or other. For example, the Salford Royal NHS foundation Trust was rated as ‘Outstanding’ in 2015, but was deemed an ‘elevated risk’ concerning ‘emergency readmissions with an overnight stay following an emergency admission’. Likewise using indicators of severe harm such as safeguarding alerts provides little help for prioritising inspections across the sector, given that the CQC received 90,000 safeguarding reports in 2014/15 alone (2015a)

Beyond these technical challenges, efforts to develop robust indicator sets have been repeatedly disrupted by institutional reforms to the sector that complicate quality assessment and regulation. While CHI was only responsible for monitoring quality in NHS organisations, the HCC and the CQC were successively responsible for monitoring quality across an increasingly vast and heterogeneous range of providers in the health and social care domains. While this increased regulatory load redoubled the CQC’s commitment to using data to assess risks to quality, it found that the expanded set of providers could not provide the same kind of quantitative outcomes-data available for the NHS. The CQC tried to make more use of qualitative sources to monitor private providers, such as online comments and complaints from users, but it struggled to devise ways to process and risk rate such information (A. Griffiths, 2012). Moreover, its generalist inspectors often lacked the expertise to interpret and use the resulting statistical analyses to guide their on-site inspection of hospitals and other complex healthcare establishments.

Even with the best data in the world, there is an irreducible normative and political element to indicator construction, selection, and application, for example related to bargains with the sector or electoral agendas. Examples include focussing on the most publically salient issues, such as access to cancer care rather than care of cardio-pulmonary conditions, the latter being nearly as great a cause
of mortality in England and Wales as cancer (Rogers, 2012). Likewise, the unevenness in indicator coverage has prompted concerns about selection bias and whether outcomes indicators can provide a sufficiently rounded and comprehensive picture of performance (Barker et al., 2004).

Moreover, the relatively small indicator sets used in the first NHS Performance Assessment Framework and by CHI for its star ratings were gamed by regulatees, who simply focused their attention on those few aspects of quality that were measured (Bevan & Hood, 2006). The CQC responded by using over 1000 quantitative and qualitative indicators, which it scored and weighted for their quality, timeliness, and fit with essential quality standards (A. Griffiths, 2012). However, the QRPs became so complicated that inspectors and regulatees alike failed to understand them. On the advice of the consultancy firm McKinsey’s, QRPs were replaced by a new ‘Intelligent Monitoring’ system, which uses a much smaller selection of unweighted indicators. As a consequence, however, the ‘proportion of Healthcare Workers with direct patient care ...vaccinated against seasonal influenza’ bizarrely has as much bearing in targeting CQC inspections as the number of ‘never events’ or MRSA incidents (CQC, 2014d).

Finally, no matter how good the underlying information, the tolerance for error in assessment is another fundamental challenge. While risk-based approaches provide a way to qualify assessments of care quality probabilistically in terms of the statistical risk of non-compliance, setting probabilistic response thresholds still involves an inevitable trade-off between type I false alarm errors and type II false reassurance errors (Demeritt et al., 2007). The Morecombe Bay scandal illustrates this dilemma: when the story went public, hospital managers and quality regulators were heavily criticised in the media for failing to spot “the first warning signs” of trouble after a baby died in a mismanaged labour in 2004 (Bunyan, 2015; cf. Chapman, Ledwith, & Borland, 2015). The official investigation report was only slightly more tolerant, noting that “Serious incidents happen in every health system because of the nature of healthcare, and no blame should be attached to staff who make mistakes”, but criticizing the Trust for “fail[ing] to identify underlying problems” in its investigation of that incident, the first in “a series of further missed opportunities to identify problems” that would lead to 12 avoidable deaths (Kirkup, 2015). The local Conservative MP also waded in, calling for the then Labour Minister for Health, Andy Burnham, to be questioned about why he had not responded more quickly and decisively. In the heat of a crisis like this, risk-based definitions of regulatory success and failure count for little if ex ante assessments of acceptable risks are ex post judged as unacceptable failures when they actually occur. Faced with public outrage and political scandal, the politician’s promise to learn the lessons and take steps to ensure that whatever happened “must never be allowed to happen again” (NHS England, 2015) sits uneasily with risk-based rationales of regulatory action.

5. **Enforcing Quality**

Efforts to calibrate regulatory enforcement in terms of risk have also proven difficult. Risk-based enforcement is conventionally conceived as a ‘pyramid’ of enforcement responses whose stringency rises in proportion to the scale of the potential harms they are intended to remedy (Ayres and Braithwaite 1992). Since the CQC recognises that “it is not feasible or proportionate to follow up every single breach of standards” (CQC, 2015b), it uses a formal risk matrix to assess the probability and consequences of regulatory breaches to “select the appropriate enforcement power”. These powers currently range from simple advice or public disclosure as strategies for encouraging
compliance to more punitive civil enforcement measures, such as warning and requirement notices to compel improvement, and in extremis criminal penalties, prosecutions and cancellation of registration (CQC, 2015c). In theory this escalating range of sanctions can then be matched to the competence, capacity and willingness of regulatees to meet regulatory requirements.

In practice, however, the CQC’s risk-based enforcement toolkit has struggled to overcome the inbuilt resistance to change that has frustrated so many other NHS reforms over the years. While serious breaches of care quality, like those found at the Mid-Staffordshire Trust, are supposed to meet with the most serious enforcement sanctions, officials are often hesitant to deploy them. The problem is not one of detection but of response, as a senior civil servant from the DoH explained to us:

“After the Mid Staffordshire disaster, there was a view that mechanisms for monitoring risks in terms of quality were not good; [having] failed to both identify and ... address the failures that were going on. Actually the main problem was addressing the failures because problems had been identified previously. There had been previous reports, including from the Healthcare Commission that had identified those problems, but nobody really reacted.”

Historically, regulators have been hesitant to use punitive civil and criminal enforcement sanctions because they can undermine the continuity of healthcare provision to the public. Unlike small care homes, hospital, ambulance and mental health trusts are often local monopoly providers, making the ultimate sanction of closure an unrealistic threat, especially for sensitive services such as accident and emergency care. GP surgeries are sufficiently numerous that the threat of closure has more deterrence value, though like hospitals, some are local monopoly providers, making their closure often politically controversial (Haynes, Lovett, & Sünnenberg, 2003; Kirkintilloch Herald, 2014; Liverpool Echo, 2015). Likewise, financial penalties can be self-defeating if they exacerbate any under-resourcing that contributed to care quality failures in the first place. This problem has been recognised and the 2013 Health and Social Care Act gave the CQC power to have trusts put into ‘special measures’, which can involve disqualification of board members, imposition of new management and forced mergers with more successful Trusts. Within two years, some 21 NHS hospital trusts, or more than 10% of all acute trusts in England, had been put into special measures. But with only a limited number of nearby trusts available to support or merge with those in special measures, there are clearly limits to this last resort approach.

More compliance-oriented enforcement tools for incentivising behaviour change in the NHS have not been particularly effective either, despite their popularity among policy-makers. Until recently, the CQC was unable to offer much education and advice because its generalist inspectorate simply did not have the expertise to do so. Moreover, any improvement notices issued by inspectors were liable to be ignored, because, as a former Special Advisor confessed to us, inspectors “had no credibility with the doctors”. This might have changed in 2013 when the CQC’s inspectorate was reformed with a dedicated expert hospital inspectorate working in specialist teams of up to thirty or more. It is not clear, however, that ignorance of what to do is the major cause of poor quality care in NHS hospitals, or for that matter ambulance and mental health trusts either. Hospitals are not obviously like restaurants where expert food safety inspectors can improve compliance by educating

---

1 Based on number of acute trusts in special measures (see: [http://www.nhs.uk/nhsengland/specialmeasures/pages/about-special-measures.aspx](http://www.nhs.uk/nhsengland/specialmeasures/pages/about-special-measures.aspx)) divided by the existing 155 acute trusts (see: [http://www.nhsconfed.org/resources/key-statistics-on-the-nhs](http://www.nhsconfed.org/resources/key-statistics-on-the-nhs)).
frequently ignorant regulatees about how to meet a narrow set of technical regulatory standards on good food hygiene (Yapp & Fairman, 2006). There is no shortage of technical expertise in hospitals; rather it is their size and complexity that makes it difficult to ensure consistent quality across the wide range of services they provide.

Likewise, naming and shaming mechanisms, such as publication of quality ratings, have created little demand-side pressure to drive up quality standards (Laverty et al., 2012). While the Ofsted-style star ratings published by the HCC and CHI from 2001 to 2005, like the Annual Health Check that succeeded them, were relatively easy to understand, they rated entire Trusts and, therefore, provided little help to patients selecting the best hospital for hip replacements. Even some erstwhile advocates of patient choice concede that point. As one member of the shadow health team put it to us, “The idea that my hospital is either safe or not safe is just not real; health and social care is so big, so diverse”.

Having been ordered by the Secretary of State to abandon the Annual Health Check because of the limited value of single-word summary judgments of hospital quality (West, 2010), the CQC tried to provide more granular information to inform patient choice by publishing inspection findings for 16 different quality standards. However, the reports were difficult for the public to understand because of the sheer number of standards and the focus on assessing compliance with minimum requirements rather than judging the overall quality of care (Nuffield Trust, 2013). Now the Government wants to go back to a more simplified summary judgement “that patients and the public can understand- driving organisations to excel rather than just cover the basics” (DoH, 2012). Accordingly the CQC is now publishing consolidated inspection grades and risk ratings for hospital trusts, GP practices and mental health services, but there is little reason to believe that patients will be able to use that information to select the best provider.

Ratings have also faced fierce sectorial resistance from medical professionals. Describing the construction of CHI’s first star rating system, a former DoH advisor recalled to us,

“When we started this, we ... thought we were just adding another piece of bureaucracy, but actually it was completely different. Inspecting one of the best heart surgeons in the world is different from inspecting a primary school teacher. There was a lot of ... anger [and] noise in the system. Instead of recognising the interest of CHI, people were rather ‘how dare you?’”

But just as the Labour Prime Minister Callaghan in 1976 famously questioned whose interests the ‘secret garden’ of educational professionals was serving, so healthcare professionals have found it increasingly hard to keep the doors closed to their own private Eden. For example, the Society for Cardiac Surgery began publishing its own risk-adjusted mortality rates for cardiac surgeons in 2007 in an effort to pre-empt the sort of externally imposed rating schemes that NHS England began publishing for other surgical specialisms in 2013. Equally, however, the CQC’s experience with publishing risk ratings for GPs underlines the difficulties in using information disclosure to nudge behaviour change and enforce quality standards. In 2015, in response to fierce complaints from GPs about the accuracy and effects on public confidence of its risk bands (Millett, 2015), the CQC agreed to publish only the raw statistics from its Intelligent Monitoring system and to “change the language used to highlight variation between practices” (CQC, 2015e).

Having NHS services publicly named and shamed is also politically sensitive for governments, who have to explain why they have allowed services to fail (Pollitt, 2010). Brandeis famously declared
sunlight to be the best disinfectant, but in the case of the NHS, disclosure of under-performance is also akin to Tsoukas’ (1997) ‘tyranny of light’, creating significant political risks for governments despite their continued efforts to shift accountability away. Although ministers have frequently likened care quality regulators to Ofsted, they have not always been as keen on seeing them act as aggressively, because “the political risk of naming a hospital as ‘failing’ was far greater than naming a school as failing” (Nuffield Trust, 2013, p. 15). As former CQC chair Baronness Young testified to the Francis Inquiry:

“the model of an independent regulator, regulating services provided by a Government Minister, was never going to be a satisfactory model. It was always going to be incredibly fraught, because inevitably both the Department and Ministers were torn between wanting good, strong independent regulation of healthcare and knowing that if good, strong independent regulation of healthcare happened, from time to time they would be put in the dock and found wanting” (Francis 2013: 941).

Other incentive mechanisms such as the principle of ‘earned autonomy’, embodied by the creation of Foundation Trust status, have not driven quality improvements. Trust status was originally framed as a reward for those NHS organisations ranked three stars or excellent by the quality regulator, offering extra-budget and increased freedom and autonomy. The new status was not accompanied by a lessening of any regulatory load, however, as might be expected from a risk-based enforcement pyramid. Even in its early light-touch days, the CQC undertook bi-annual inspections of every hospital every two years, becoming annual inspections after the Winterbourne View scandal in 2011, and relaxing back to 2 years in 2013. In fact, the 2008 decision by the DoH to decouple the link between excellence and autonomy by seeking to transform all providers into foundation trusts, simply put pressure on care quality. As a senior DoH civil servant explained to us:

“At the time Brown declared that he wanted all NHS trusts to become foundation trusts, this became the new priority of the department. And the regulators had to make many decisions regarding borderline cases. We knew that Midstaff was one of these. From that moment, most of the NHS managers’ goals became to achieve this status of FT, they were pressed to do that, their career depended on that, at the expense of other goals”.

Mid Staffordshire hospital trust was part of the 2008 wave of applications and, as the Francis report (2013) highlights, the requirement to balance the books and demonstrate value for money resulted in substantial cuts in staff and savings in medical procedures in an organisation that had been widely known to be under stress since the 1980s. Risk-based methods for enforcing quality are likely to be of little use in the face of such perverse incentives.

6. Discussion and conclusions

Risk-based regulation has been a central dogma of regulatory reform in the UK for many years now, and in some domains, such as occupational health and safety, it has won widespread support as a way of optimising regulatory goals and efficiently allocating monitoring and enforcement resources (Demeritt et al 2015). Yet the analysis above suggests that risk instruments have failed to improve the regulation of healthcare quality, not least because healthcare quality itself has proved fundamentally hard to regulate for at least three reasons.
First, the goal of healthcare quality regulation has been unstable. Unlike occupational health and safety where it has long been accepted that the regulatory goal should be a reasonably practicable trade-off between safety and cost, there has never been any lasting consensus amongst professionals, publics and governments about what healthcare quality actually is and what standards of healthcare would be acceptable. As a result, definitions have constantly shifted over time, becoming little more than a laundry list of complex and often competing goals that are simply layered on top of each other according to changing fashion and political whim. If quality regulation in the NHS has failed, it is at least partly because quality itself and the goals of regulation cannot be defined.

Second, despite access to what is probably the largest database on outcomes and performance for any regulated sector in the world, NHS regulators have also struggled to measure the quality of care and to identify which providers are failing to deliver healthcare adequately. Quality assessment failures have been multiple. Reasons include difficulties in: capturing outcomes for patients and “what matters” to them; making credible inspection judgements about complex healthcare organisations; interpreting vast quantities of heterogeneous and conflicting data; and adapting measurement methods and indicators to frequent changes in policy and the organization of healthcare.

Third, regulators have only had limited tools for improving quality and enforcing standards. Deterrence oriented sanctions for dealing with noncompliance risk compromising the quality and continuity of healthcare delivery. More compliance oriented levers of behaviour change such as earned autonomy, education and advice, or public disclosure, however, can create perverse incentives, alienate expert clinical staff, or fail to supply sufficiently granular information about performance to be useful in leveraging change in a sector characteristically dominated by monopolistic provision.

If we use the metaphor of a control system to conceptualise regulatory regimes, then a prerequisite for their success is that regimes must have clear goals and that they can monitor- and secure changes in- regulatee behaviour to ensure compliance with those goals (Hood et al., 2001). Our analysis, however, shows that healthcare quality regulation has struggled on all three counts: acceptable quality standards are hard to define, difficult to assess, and impossible for regulators to enforce completely.

While risk-based approaches to regulation appear superficially attractive as a way of coming to terms with the inevitable limits of what regulation can hope to achieve, they have not solved the problems facing healthcare quality. The failure of risk-based regulatory approaches in this domain is strikingly illustrated by the CQC’s laundry list of vague regulatory goals and ambiguous quality standards: safe, effective, caring, responsive and well-led. Such lists may have a feel-good factor about them but if a key risk to reaching one goal- such as ‘safety’- is compliance with another goal- such as the ‘effective use of resources’- then risk-based solutions will only work if it is clear how to identify, measure and manage such trade-offs. Rather than confronting these questions about care quality regulation, different goals have simply been layered on top of each other with little thought of how to define, assess, or ensure the optimal balance between them. As a consequence, regulators and the NHS have struggled to distinguish between unacceptable failures and the acceptable risks of healthcare. Even when they have agreed ex ante on what risks are in principle acceptable, political
sensitivities and public outrage mean that adverse outcomes are rarely regarded as acceptable when they come to light.

Setting aside these challenges of defining and assessing risk-based quality goals, it is not clear that risk-based prioritisation of regulatory enforcement will do much to improve healthcare quality in the NHS either. The fact that more than 80% of the acute trusts inspected in England over the last two years have been found ‘inadequate’ or to ‘require improvement’ suggests that the challenge is not one of targeting the odd ‘bad apple’ for improvement. Rather, if the inspection findings are to be believed, then the problems affecting healthcare quality in the NHS are widespread and endemic. With failures to be found seemingly wherever inspectors look, then efforts to improve risk-based targeting of enforcement activity rather miss the point about the scale of the problem and of the measures required to ensure quality in the NHS.

The challenges facing risk-based regulation have been made worse by seemingly constant reorganisations both of the sector itself and its regulators, leading to new problems and further reforms in response. Consequently, regulation has tended to be a ‘tombstone’ to the last disaster, ironically often accompanied by the recycling of previously discredited ideas. Thus Ofsted-style ratings were introduced in 2001 to drive-up standards and increase patient choice, then abolished for being misleadingly over-simplistic, only to be reintroduced some years later. Likewise, expert regulators for CHI led teams of peer reviewers on in-depth inspections, only to be replaced by the CQC’s generalist inspectorate that conducted light-touch inspections, but that approach was then discredited and dropped in favour of in-depth comprehensive inspections led by specialists assisted by clinicians and ‘experts by experience’. As one senior DoH official put it to us:

“The pendulum swings with regulation between ... targeted or risk-based, and between announced and unannounced. I mean there are a lot of choices about how you design an inspection regulation organisation, and frankly, we haven't got any evidence ... Things were very well documented about what wasn’t working. But as I say, I'm not quite sure that there's evidence to really say that another way of doing it is really better.”

As a consequence, regulatory models and institutions rarely have long to bed down before they themselves or their regulatees are disruptively reformed once more. Thus developing a sensible and stable priority planning system and recruiting, developing and maintaining an inspectorate with the right skills to inspect a constantly changing set of providers in a ‘risk-based’ way has been a real challenge.

Such arguments suggest that the problems confronting healthcare regulation go far deeper than questions of optimizing regulatory efforts according to risk. Rather there are genuine questions about what can realistically be expected from regulation in terms of improving care quality and ultimately meeting public expectations (Dixon, 2012; Torjesen, 2012). If the problems facing healthcare quality are systemic problems of delivering a service in the context of numerous conflicting constraints, rather than unwitting or intentional non-compliance with regulatory rules, then is healthcare quality really something that admits of regulatory solutions? As a member of the shadow health team put it to us:

“Maybe we need to be realistic about what could be achieved through regulation. I like the new inspection model developed by the CQC, but it is probably going to fail again within the next two years.”
Until there is greater consensus about what care quality means and how far it is reasonable to go to ensure that goal, then risk-based approaches to healthcare regulation are unlikely to offer much help in fixing the problems facing the NHS.

Acknowledgements

Research for this paper was supported by a grant awarded through the Open Research Area Programme for the Social Sciences from the ESRC (ES/K006169/1) (HowSAFE: How States Account for Failure in Europe: Risk and the Limits of Governance), which funded Beausissier, Demeritt and Rothstein. While we are grateful to our informants for sharing their time and insights with us, the authors alone are responsible for any errors of fact or analysis herein, which does not necessarily represent the views of the Care Quality Commission, which employs Griffiths.

References


Griffiths, A. (2012). *In Theory & in Practice: Can Risk-Based Approaches to Regulation Work in the Care Quality Domain?* (MSc.). King’s College London, London.


