



# Contracting for Quality in the NHS

Putting the Francis Report  
in Perspective

Contracting for Quality in the NHS: Putting the Francis Report in Perspective

20th Annual Lecture

Professor Alan Maynard

16 July 2013

Royal College of Physicians

London

*About the Office of Health Economics*

Founded in 1962, the OHE's terms of reference are to:

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

The Office of Health Economics receives research and consulting funding from many sources, including the Association of the British Pharmaceutical Industry (ABPI).

*The views expressed in this publication are those of the author and do not necessarily represent those of ABPI or OHE.*

Office of Health Economics  
Southside, 7th Floor  
105 Victoria Street  
London SW1E 6QT  
United Kingdom  
[www.ohe.org](http://www.ohe.org)

©2014

All rights reserved

Printed in the United Kingdom

ISBN 978-1-899040-49-0

## **About the author**

Professor Alan Maynard began his career at the University of Exeter (1968–1971) as an assistant lecturer and lecturer in Economics. Between 1971 and 1977, he lectured in Economics in the Department of Economics and Related Studies, University of York, where he later was Founding Director of the Graduate Course in Health Economics and Founding Director of the Centre for Health Economics. He currently is Professor of Health Economics in the Department of Health Sciences at the University of York.

Professor Maynard has been involved in NHS management since 1983, initially as a non-executive director and subsequently as Chair of the York Hospitals NHS Foundation Trust (1997-2010). Currently, he is Lay Chair of the Vale of York Clinical Commissioning Group.

Professor Maynard's work extends well beyond the UK: he has worked as a consultant for the World Bank, the World Health Organisation, the European Union and the UK Department for International Development in two dozen countries.

A Fellow of the Academy of Medical Sciences, Professor Maynard also has honorary doctorates from the Universities of Aberdeen and Northumbria. He was awarded the OBE for services to the NHS in 2009.

## **Contents**

Introduction .....	1
The NHS: Principles and Practice .....	1
Patient safety .....	1
Unwarranted clinical practice variation .....	2
NHS Policy Responses to Problems of Clinical Practice .....	3
An Economic Approach to Health Care Reform .....	4
The principal–agent relationship .....	4
Outcome measurement: a long, slow journey .....	5
Incentivising improved practice .....	7
The role of financial incentives .....	7
The role of competition .....	8
The role of non-financial incentives .....	8
Conclusions .....	9
References .....	11

## Introduction

When scandals or crises hit the National Health Service (NHS), politicians tend to mitigate their distress by creating a public inquiry. Thus, deficient medical practice by a paediatric cardiac surgeon in Bristol, along with a number of other clinical failures over the years, resulted in such public inquiries. These inquiries, usually chaired by lawyers and focused on regulation, often lack systematic consideration of research evidence and their findings tend to be implemented in theory, rather than resulting in development of observably efficient practice.

The poor performance of the Mid Staffordshire NHS Foundation Trust led to a public inquiry chaired by Robert Francis (Francis, 2013). His three-volume, 1,794-page report had an executive summary that itself was 25 pages. Francis made 290 recommendations for more regulation. Sadly, his proposals were not evidenced, not prioritised and not costed.

There are better ways to investigate and improve patient care, in particular by the use of evidence and economic analysis. It is exceptional that such an approach is used to improve the NHS, with the Merrison Royal Commission of 1979 being a rare example. The use of an evidenced approach recognises that all public and private health care markets are inefficient and inequitable. Regulation is therefore inevitable. But the alternative to *more* regulation (as advocated by lawyers and public inquiries) is *better* regulation. Insights from economics can contribute to the production of improved regulation that protects both taxpayers and patients.

## The NHS: Principles and Practice

The introduction of the NHS created income protection in times of catastrophic illness for all citizens regardless of their willingness and ability to pay. It continues to provide access to care on the basis of patient need. This requires careful interpretation and, in order to be implemented, clear evidence of the relative cost effectiveness of different treatment options (Williams, 1978). The NHS, tax funded and mostly free at the point of delivery, offers government the necessary conditions for expenditure control and equity of access to care. These are attributes of a civilised country that should create both pride and a sense of security for UK citizens.

However, the NHS has deficiencies in clinical practice. These are largely identical to those of public and private systems internationally, but are often reported parochially as unique to the NHS.

## Patient safety

Current media coverage seems to imply that health care, especially hospital provision, should be completely safe. This is manifest nonsense. A US institute of Medicine publication argued years ago that “to err is human” (Kohn, et al., 2000). Health professionals make mistakes and patients are damaged. The consensus, supported by the World Health Organisation, is that ten per cent of patients are routinely damaged by hospitalisation. Error rates in primary and community care are largely unknown.

The litany of “scandals”, their associated mortality and morbidity, and the recommendations of public inquiries have produced all too little learning and improved patient protection. The Bristol Inquiry into paediatric cardiac mortality, for example, showed institutional awareness of avoidable death and damage to children, and failure to act on available data. Finally, an investigation was initiated by an anaesthesiologist who, because of it, subsequently was forced to migrate to Australia to find a job. The public inquiry linked to Bristol cardiac surgery identified the need for centralisation of facilities, which has been only partly implemented, and it may also have improved transparency in surgery outcomes, an area in which the NHS now leads the world.

In Bristol, and with other high-profile clinical cases such as those of gynaecologists Ledward and Neale and the serial killer GP Shipman, retrospective analysis of routine NHS data showed them to be outliers

who should have been investigated and prevented from continuing to practice much earlier than they were (Harley, et al., 2005).

## **Unwarranted clinical practice variation**

Unwarranted variation in health care delivery is manifested in all aspects of clinical practice. This issue has been evident for decades, but remains unresolved due both to debate about its magnitude and knowledge of efficient remedies.

A medical officer in the Ministry of Health, Glover, was a pioneer in pointing out variation. In 1938, he published the results of his analysis of the more than doubling of tonsillectomy rates in England and Wales between 1923 and 1931 as reported in school medical service data. He traced the steady increase in this activity in the mid-1930s, particularly in public schools where parents could pay for treatment. He estimated that 85 children under 15 years of age died each year after the procedure and quoted a Medical Research Council conclusion that “there is a tendency for the operation to be performed as a routine prophylactic ritual for no particular reason and with no particular result” (Glover, 1938, p. 14).

Nearly 40 years later, Bloor and colleagues reported similar variations in tonsillectomy rates in two regions of Scotland (Bloor, et al., 1977). One region practised “watchful waiting” while the other responded to immediate patient characteristics. Evidence of the appropriateness of the latter approach was absent.

This British material is dwarfed by the work of Wennberg and his colleagues at Dartmouth College in the US that began in the 1970s (Wennberg, 2010). This research uses US Medicare data and analyses variation in clinical practice. It challenges the “more is better” ethos in health care and asserts that cost savings of 30–40 per cent of the Medicare budget could be made if “conservative safe practice” was delivered uniformly across the country (Fisher, 2003; Wennberg, 2010).

Debate continues about the magnitude of savings that could be made by reducing clinical practice variation. Some US economists using longitudinal analysis estimate savings of 12–15 per cent of Medicare spending (Cutler and Sheiner, 1999; Rettenmeier and Wang, 2013). One critique asserted that the Dartmouth methods used units of analysis that do not adequately capture variations in need that are associated with social class (Cooper, 2013). Sheiner has controversially asserted that “variations in health care spending do not provide a useful measure of the inefficiencies of our health care system” (Sheiner, 2013).

The potential of addressing variations in health care spending has been effectively underlined by an Institute of Medicine review of the US debate that is subtitled “target decision making, not geography” (Newhouse, et al., 2013). This concludes that significant unwarranted variations exist in public and private health care in the US. In the private sector, variations appears to be caused primarily by variations in the price of health care, with ability to pay determining use. Under publicly funded Medicare, where prices are largely fixed, variations are due to volume differences as greater use creates more income for providers. The difference in Medicare are largely in post-acute care. Newhouse and Garber (2013), for example, note that an estimated 73 per cent of variation would be eradicated under Medicare if post-acute differences in patient care were eliminated.

One aspect of the variations debate is the potential for “over-diagnosis” in clinical care. How many patients must be treated with “preventive” care to save a life? What is the appropriate number of patients that must be treated to save one patient-life in the populations that appear to have, for instance, “high” blood pressure or “high” levels of cholesterol? In an effort to save lives, clinical guidance has progressively advocated intervention with pharmaceuticals for lower and lower levels of blood pressure, cholesterol, diabetes and osteoporosis. It has been estimated that altering the diagnostic criteria for hypertension in the US in 1997 increased the number of cases by 35 per cent.

Reducing the fasting sugar levels for the diagnosis of diabetes produced 1,681,000 new cases (Welch, et al., 2011).

Convincing practitioners to follow such guidance may produce profits for the health care providers, but its efficiency is in doubt: the cost of saving one life with these inflated intervention criteria is high, and diagnosing patients as “at risk” may reduce their quality of life unnecessarily.

## **NHS Policy Responses to Problems of Clinical Practice**

In the UK and elsewhere, the problems of patient safety, clinical practice variation and over-diagnosis are continually revisited in policy documents and research. The research merry-go-round is replicated in policy debates that often clearly identify the problems, but fail to devise evidenced methods of protecting patients and taxpayers.

Emulating the international policy debate about clinical practice variation, UK policy has identified provider problems for decades. Thus, in 1976, the government published a document entitled, *Priorities for Health and Personal Social Services*. This identified the scope for savings if variation could be reduced—for example, £40 million would be saved if all hospitals could emulate the lengths of stay of the best quartile (DHSS, 1976).

More recently, the Labour government commissioned the consulting firm McKinsey & Company to advise on how to achieve savings (McKinsey & Company, 2009). Their findings identified a potential £20 billion in savings that the coalition government then adopted: the so-called “Nicholson challenge”, requiring four per cent annual NHS savings for four years. The McKinsey report was based on differences in activity and resource, and its findings were predicated on the belief achieving average performance by all hospital providers would create savings of £20 billion.

The 1976 government report, the McKinsey report, and other reports related to variation all asserted that large resource savings would be achieved if differences could be reduced by reducing dispersion and shifting average performance. None of these reports articulated how such change could be achieved cost effectively. Thus, policy has continued to recognise the problem, but fought shy of identifying and implementing efficient ways of reducing waste (Maynard, 2013).

A former President of the Royal College of Physicians, Sir Raymond Hoffenburg, criticised the Thatcher reforms of the NHS by declaring: “Instead of ready, take aim and fire, the Government chose to make ready, fire and then take aim!” (quoted in Maynard and Bloor, 2003, p. 535). This is an apt description of the continuous “re-disorganisation” of NHS structures over 40 years.

Edwards lists over 70 minor and major “reforms” of the NHS since 1974, when structural change first began (Edwards, 2010). After this first restructuring, the Merrison Royal Commission criticised the changes and concluded that reorganisation had created:

. . . an immense amount of administrative work in the preparation for new machinery; disruption of ordinary work, both before and after reorganisation caused by the need to prepare for and implement the changes; the breakdown of well-established formal and informal networks; the loss of experienced staff through retirement and resignation; the stress and strains of some staff having to compete for new jobs (quoted in DHSS, Great Britain Welsh Office, 1979, p. 5).

Despite these criticisms, NHS reform continued with Prime Minister Thatcher’s “marketization” and the creation of the purchaser–provider split; Prime Minister Blair’s initial abandonment of GP fundholding, and then its resurrection with Primary Care Groups and Primary Care Trusts; large increases in funding with regulation; and faith in “patient choice” as the driver of change. The Health and Social Care Act of 2012 imposed yet another change in structure and hoped for improvements in patient care.

This continued process of “re-disorganisation” demonstrates a failure to evaluate change systematically and a willingness to impose large opportunity costs arising from reform—for example, the recent Lansley reforms cost £3 billion (Bradshaw, et al., 2013) with little evidence of benefit.

How has this happened? One cause of this waste has been political rhetoric. In his discussion of the “ideological jungle”, Williams (1988) identified two competing ideologies, the libertarian and the egalitarian. Adherents to both views tend to deploy the ideal version of their preferred options and the real version of their opponents’ system. Thus, if a market orientated Conservative compares the theoretical virtues of private systems with the deficiencies of the actual NHS, a Labour follower takes the ideal public health system and compares it with the deficiencies of actual private health care. Thus, political exchanges tend towards mindless blame games, failing to recognise that both public and private systems fail for similar reasons.

Another cause of often-fruitless NHS reform is the superficial nature of change processes and their failure to use and enhance the evidence base about how to improve resource allocation and patient care. What Campbell noted over 40 years ago remains dominant:

It is one of the most characteristic aspects of the present situation that specific reforms are advocated as though they were certain to be successful. For this reason, knowing outcomes has political implications . . . Ambiguity, lack of truly comparable comparison bases, and lack of concrete evidence all work to increase the administrator’s control over what gets said, or at least to reduce the bite of criticism in the case of actual failure. There is safety under the cloak of ignorance (Campbell, 1969, p.2).

The combined forces of ideological values, rather than evidence-informed reform, and policy makers’ evasiveness in failing to design and measure costs and effects to protect their reputations, has consumed significant resources, but with little benefit to patients and taxpayers. Can a more economic approach do better?

## **An Economic Approach to Health Care Reform**

### **The principal–agent relationship**

Health care is characterised by information asymmetries. This complicates contractual relationships among patients, doctors and health care organisations. As Arrow argued, “Contractual relations are frequently a good deal more complicated than the simple models of exchange of commodities and services at fixed prices would suggest” (Arrow, 1975, p. 48).

There is an information asymmetry between medical experts, such as doctors, and the patient. For example, as a cancer patient, I trust my physician-haematologist to do his best to control my disease. My understanding of the technicalities of diagnosis and treatment is limited. I trust him to strive to maintain my health status as I struggle to understand the disease and its effects on my co-morbidities.

Information asymmetry also occurs between the purchaser (government, Clinical Commissioning Groups (CCGs)/private insurers) and providers. Not only are CCGs price-takers because of national tariffs, they are also quality-takers because they are ill informed by routine data about patient outcomes and complication rates.

At the level of the hospital, information asymmetry exists between managers and clinicians. The former have limited information about cost and outcome data. They are driven by the imperative of expenditure control and financial balance, taking quality and cost largely on trust.

Sadly, information asymmetries may exist across doctors, even when they are part of the same team and their directorates depend on good team performance.

The common problem is that principals do not know whether their agents are doing their best for them (Arrow, 1963). Furthermore, in health care the agent/doctor may have more than one principal for which he is acting. Blomqvist (1991) described doctors as being “double agents” acting on behalf of both their employers and patients. In the hospital sector particularly, the relationship between doctors and their employers is characterised by contracting between two firms. Their interactions may or may not ensure the delivery of safe and efficient patient care (Harris, 1977).

This complex network of relationships between principals and agents lacks information to inform efficient choices. Consequently, most transactions in health care markets are based on trust: patients trust doctors to do their best for them, and purchasing agencies, such as CCGs and private insurers, trust providers to provide good quality care.

The failure of doctors and other providers to ensure the delivery of high quality care has been identified with increasing frequency in recent decades. Policy makers have responded with regulations to minimise aspects of care such as “never events” and “avoidable mortality”. They have sought to enforce their will through fines and bonuses. However, the opaqueness of the “medical monopoly”, in terms of the activity and outcomes of doctors, has remained largely unchallenged by policy makers who are either too ignorant or are too intimidated by the power of the medical profession.

## **Outcome measurement: a long, slow journey**

The lack of transparency in contracting between public and private purchasers and providers in the health care market produced by the principal–agent relationship is epitomised by slow progress in addressing the issue of patient outcomes. Policy makers have tended to focus reform efforts on structural change and process measures of “success” such as reduced waiting times. They have ignored a long history of advocacy of outcome measurement by members of their own profession.

In the 1790s, conflict erupted between doctors at Manchester Royal Infirmary and their employers. To deal with an outbreak of typhus, the hospital hired more doctors, which led those already in post to resign. Percival was asked to mediate; one consequence of this was the publication of his book *Medical Ethics* (Percival, 1803). In this book, he advocated both outcome measurement and incentives for improvements in clinical practice. He proposed that outcomes be measured in terms of whether patients were “cured, relieved, discharged or dead”. Percival argued that with such data, “Physicians and surgeons would obtain a clearer insight into the comparative success of hospitals and private practice; and would be incited to diligent investigations of the causes of such difference” (Percival, 1803, p. 16).

Over two hundred years later, health care systems remain tardy in measuring patient outcomes and using comparative data to “incite” changes in clinical behaviour. Percival’s insights were replicated throughout subsequent years.

The UK Lunacy Act of 1845 required mental hospitals to measure outcomes in terms of whether patients were dead, recovered, relieved or not improved. Failure to collect this information carried a fine of £2, an early example of pay-for-performance incentives.

Florence Nightingale further reformed the outcome measures of Percival and the Lunacy Act by advocating the measurement of outcomes in terms of whether patients were dead, relieved or unrelieved. She argued that outcome measurement was essential:

I am fain to sum up with an urgent appeal for adopting this or some *uniform* system of publishing the statistical records of hospitals]. . . .

. . . . In attempting to arrive at the truth, I have applied everywhere for information, but in scarcely an instance have I been able to obtain hospital records for for any purpose of comparison. If they could be obtained, they would enable us to decide many other questions besides the ones alluded to. They would show subscribers how their money

was being spent, what amount of good was really being done with it, or whether the money was doing mischief rather than good . . . (Nightingale, 1863, pp. 175-176)

While the Lunacy Act system of outcome measurement survived in the psychiatric sector until the creation of the NHS in 1949, some acute hospitals that adopted similar systems carried on classifying patients by outcome until later; the London Hospital, for example, abandoned the system only in 1968.

Similar debates about outcome measurement took place in the US. Codman advocated measurement of the "end results" for his patients, that is, systematic follow-up after surgery. His advocacy resulted in his losing his practicing rights at Massachusetts General Hospital. He noted that "It is against the individual interests of the medical and surgical staffs of hospitals to follow up, compare, analyse and standardise their results" (Codman, 1917, p. 8).

Codman set up End Results Hospitals, but the effort failed. Although he published his outcome data, Massachusetts General did not. Instead, it successfully asserted its results were better and convinced the public they were. The lessons learned by Codman are relevant today: for transparency of outcomes to be beneficial, clinical engagement and universal participation are essential.

Repetition of the advocacy of measurement and management of patient outcomes continues. For instance, US President Reagan ordered the publication of mortality data by hospital in 1983. Two American economists, Dranove and Satterthwaite (1992), used cardiac data from New York State and the Commonwealth of Pennsylvania to show how problems of risk-rating mortality data could lead to patient selection that denied needed surgical intervention to patients with complex needs.

Since 1985, the UK National Confidential Enquiry into Patient Outcomes (NCEPOD) has been used to collect surgical complication and mortality data. This was confidential and voluntary; one of its authors claimed that his involvement in NCEPOD delayed his acquisition of a distinction ward (bonus) for five years due to peer opposition to his work.

More recently, Keogh and his colleagues at the Society for Cardiothoracic Surgery in Great Britain and Ireland have collected and reported complication and mortality data for all their members who are specialist surgeons. They have shown that transparency of comparative data has "incited" the change that Percival believed it would. Average performance has improved and variation has declined. They claim there has been a 50 per cent reduction in risk-adjusted mortality (Bridgewater, et al., 2011). In his role as Medical Director of NHS England, Keogh has now required ten surgical specialties to publish similar data and is extending this to some physician services. Progress has been slow since Percival's book in 1803, but perhaps outcome measurement at the individual practitioner level is arriving at last.

The collection of comparative mortality and complication data is useful, but an essential supplement is the effect of medicine and surgery on the patient's quality of life (Kind and Williams, 2004). Patient-reported outcome measures (PROMs) were introduced in the English NHS in 2009. Using validated generic (EQ-5D) and condition-specific (for example, the Oxford hip score) quality-of-life measures, patients' physical and psychological functioning before and after health care are tracked to determine any changes in quality of life. Initially, PROMs were used for four surgical interventions: hip and knee replacements, hernia repair and varicose vein surgery. Significant improvement in the quality of life of patients after knee and hip replacements was demonstrated, but nearly half of hernia patients reported poor quality of life: the latter may support the case for more watchful waiting. Further work has been carried out on six chronic disease categories and for some coronary artery bypass surgeries.

The medical profession has been continually encouraged to collect and use outcome data to incite change. Until recently, however, it has manifestly failed to respond or to protect patients from poor practice. However, Keogh's activities to improve transparency, along with the collection of patient-reported outcomes, suggest that change may at last be happening.

## **Incentivising improved practice**

Given the existence over many decades of significant variations in expenditure, activity and outcomes, the primary policy focus now is on how to produce change in clinical practice. The current Pavlovian response of policy makers is that “incentives” need to change. But which incentives, and how? Should we “pay ‘em” with financial rewards and penalties or “flay ‘em” with comparative data that impugns their reputation and, as Percival hoped, “incites” them to change? As ever in health policy, these are contentious issues informed by evidence that itself is the subject of debate.

### **The role of financial incentives**

The current focus of the incentives debate is how to deploy financial rewards and penalties to alter behaviour, and whether these should be used at the hospital or clinician level. Two examples of this are the incentivisation of UK GPs with the “quality and outcomes framework” (QOF) and the Premier-Medicare programme to incentivise hospitals in the US, translated into the “Advancing Quality” programme in northwest England.

The UK QOF awarded points that, when earned for particular activities, translated into enhanced pay for teams of primary care physicians. These activities were loosely evidence-based initially (Fleetcroft and Cookson, 2006), were achieved rapidly with some evidence that target levels were already being achieved before QOF (Serumaga, et al., 2011) and at a cost exceeding £1 billion. Evidence showed that service delivery was improved and that the initial laggards changed their practices significantly (Doran, et al., 2008).

The US Premier-Medicare hospital programme involved a group of hospitals interested in applying financial incentives that volunteered to participate (CMS, 2013). The programme incentivised five clinical areas: myocardial infarctions, heart failure, hip and knee replacements, pneumonia and coronary heart bypass surgery. For each condition, the Centers for Medicare & Medicaid Services (CMS) selected a series of process performance measures (e.g. use of aspirin and beta-blockers after heart attacks) and an outcome measure, mortality.

Each participating hospital measured its initial achievements against these targets. The top decile of performers was rewarded with a two per cent bonus on Medicare payment and the second decile received a one per cent bonus for subsequent performance. Penalties of two and one per cent respectively were proposed for the lowest and second-lowest deciles unless their performance improved. Performance of these US hospitals improved in terms of the activity targets, that is, hospitals provided better practice in terms of process measures. Despite this, there was no observable effect on mortality outcomes (Jha, et al., 2012).

However, a mortality benefit was found as the result of English adoption of a policy similar to that used in the US (Sutton, et al., 2012). Using the same activity and outcome measures, NHS hospitals in the former North West Strategic Health Authority were incentivised with larger (four per cent) bonuses if they achieved ranking in the top decile. This mortality effect appears to demonstrate that the UK version of the programme was cost effective (Meacock, et al., 2014).

The evaluation of financial incentives to alter clinical performance is extensive and growing. However, it often is not well designed with comparators and evidence of cost effectiveness (Maynard, 2012). Furthermore, use of penalties as an incentive has been limited. Both Adam Smith and present-day behavioural economists have argued, and shown, that the threat of loss has a powerful effect on performance (Smith, 1759; Kahneman and Tversky, 1979). Thus, for instance, would paying a bonus at the start of a year with the threat of its loss if performance is poor produce more or less change than the payment of a bonus at year’s end? Smith believed that:

Pain . . . is, in almost all cases, a more pungent sensation than the opposite and correspondent pleasure. The one almost always depresses us much more below the

ordinary, or what may be called the natural state of our happiness, than the other ever raises us above it (Smith, 1759).

Payment for performance experiments such as QOF, Premier-Medicare and Advancing Quality, demonstrate that financial incentives are associated with improvements in process measures of performance and sometimes even with improved outcome in terms of reduced mortality. However, does correlation prove causation?

The financial incentive programmes involve detailed measurement of activity and outcomes, creating increased transparency of comparative performance. Is the performance change the result of this transparency and/or the financial incentives? Separating out the effects of these two elements is difficult.

### **The role of competition**

Like financial incentives, competition is often advocated as a mechanism for improving efficiency. As with the literature on financial incentives, the evidence is contentious.

The market for goods and services is characterised by variations in competitive conditions. Market entrants, such as the Japanese firms Sony and Toyota in the 1970s and Chinese-export orientated firms more recently, have exploited comparative advantage. They have destroyed European and American competitors, providing strong incentives for survivors to innovate to compete and sustain market share and profits (Syverson, 2011). Can competition be used in a health care system to improve efficiency?

Work by Gaynor, et al. (2010) and Cooper, et al. (2011) offers some evidence that non-price competition may induce greater efficiency in specific markets. Thus, in London, competition to improve care in the NHS market appears to have reduced mortality rates from heart attacks (Cooper, et al., 2011; Gaynor, et al., 2010). The mechanisms of effect, and whether such effects can be produced in more dispersed markets, are unclear.

Can private sector competition improve NHS care? The current private sector is small and specialises mostly in elective surgery such as hip and knee replacements. They appear to have no quality advantages except in hotel facilities, and have proven unwilling to compete on price. Investments by the Blair government in "independent sector treatment centres" to reduce NHS waiting times have also failed to demonstrate greater efficiency in terms of observable quality and price.

Where competition appears to have improved outcomes in the NHS, the effect may have been produced not by financial incentives, but comparative outcomes data inciting change.

### **The role of non-financial incentives**

Performance measurement and transparency in terms of comparative performance are the essence of non-financial incentives. Performance data, in particular practice variation, enables doctors and their teams to identify their location in the distribution and highlights those in the "good" and "bad" tails of the distribution. Such transparency had significant effects in the development of the Japanese car industry decades ago (Deming, 2000).

The mechanisms through which comparative data alters behaviour are related to concepts such as reputation, duty and trust. For instance, Arrow argues: "Professional responsibility is clearly enforced in good measure by a system of ethics, internalised during the education process and enforced in some measure by formal punishments and broadly by reputations" (Arrow, 1985, p. 50).

Transparency may make clinicians more defensive of their reputations and ensure they are anxious, or in Percival's language "incited", to maintain responsible activity and outcomes in performance data.

Adam Smith also expressed the importance of non-financial incentives:

Those general rules of conduct, when they are fixed in our mind by habitual reflection, are of great use in correcting the misrepresentations of self-love concerning what is proper to be done in our particular situation. . . .

The regard of those general rules of conduct, what is properly called a sense of duty, is the principal consequence in human life, and the only principle by which the bulk of mankind are capable of directing their actions (Smith, 1759).

Reputation and duty underpin another non-financial incentive: trust. Confucius argued that three things are needed to retain political power and public credence: food, weapons and trust. If a ruler cannot hold on to all three, he should give up weapons first and food next, but trust must be guarded to the end. "Without trust a people cannot stand", he argued.

When delivering the 2002 Reith Lecture, the renowned ethicist O'Neill argued: "We need [trust] because we have to be able to rely on others acting as they say they will, and because we need others to accept that we will act as we say we will" (O'Neill, 2002).

Akerlof and Kranton (2010a, 2010b) argued that "performance pay demonstrates bad faith. It tells employees that they are not trusted to do the right thing" (Akerlof and Kranton, 2010b). Like Smith over 200 years ago, they sought to emphasise the role of social norms: "People's notions of what is proper, and what is forbidden, and for whom, are fundamental to how hard they work, and how they learn, spend and save" (Akerlof and Kranton, 2010a).

How do norms and identity evolve over time? Like other decision makers, doctors can regret past norms and make choices that alter their identity and behaviour. Social norms, and associated behaviours such as clinical practice variation, are socially unacceptable and demonstrably inefficient. However, not yet clear are the relative roles of financial and non-financial incentives in improving efficiency in a cost effective manner.

## Conclusions

Public policy aimed at improving the efficiency of the NHS has been traditionally focused on structural "re-disorganisation" since 1974 (Edwards, 2010). For a quarter of a century, reformers have sought to improve the performance of the purchaser-provider market. The continuing policy theme has been that the Utopia of efficient, "patient centred" health care was achievable. But even one of the supporters of the internal market reforms is now suggesting that it is time to pause and reflect: "If one day subsequent generations find you cannot make commissioning work, then we have been barking up the wrong tree for twenty years" (Ken Clarke quoted in Timmins, 2012).

A principal problem of the NHS market is that purchasers are price and quality takers, whilst providers are price and quality makers. Francis recognised this problem:

Commissioners—not providers—should decide what they want to provide. They need to take account of what can be provided, and for that purpose will have to consult clinicians both from potential providers and elsewhere, and be willing to receive proposals, but in the end it is the commissioner whose discretion will prevail (Francis, 2013, p. 1,688).

This idyll ignores the fact that national tariffs determine prices, and that provider quality in terms of process and outcomes is opaque due to the enduring unwillingness of the medical profession to collect and systematically use comparative performance data.

Policy makers have been "jumping on the spot" and avoiding the clearly defined cause of practice variation and inefficiency—the doctor, who is the leader of health care teams and the appropriate focus of reform. However, changing medical practice is something that can be best led by the medical

### *Contracting for Quality in the NHS*

profession itself, as epitomised by Keogh's recent implementation of the advice of Percival from over 200 years ago!

It is time to pause and reflect on the failures of conventional structural reforms in the NHS since 1974. Continuous "re-disorganisations" have considerable opportunity cost and have produced little evidence of benefit for taxpayers and patients. Contracting for quality and patient protection in the NHS, and all other health care systems, requires scepticism about medical practice and rigorous performance review of clinicians by clinicians. It is remarkable that the advocacy of medical "radicals" such as Percival and Codman have had so little impact even after such a long time. Unwarranted and potentially wasteful clinical practice variations remain. Outcome measurement and transparency are recent additions to our knowledge, but apply to just a fraction of everyday care.

Clinical autonomy and professionalism remain important, but they must be accompanied by transparency. Reputation should be the engine of quality improvement that is led by clinicians, rather than repetitive political reorganisations and increases in legalistic regulation. Physicians, heal the NHS!

## References

- Akerlof, G. & Kranton, R., 2010a. *Identity economics: how our identities shape our work, wages and well-being*. Princeton, NJ: Princeton University Press.
- Akerlof, G. & Kranton, R., 2010b. It is time to treat Wall Street like Main Street. *Financial Times*, 24 February.
- Arrow, K., 1963. Uncertainty in the welfare economics of medical care. *American Economic Review*, 53 (5), pp. 941-973.
- Arrow, K., 1985. The economics of agency. In: J. Prett & R. Zeckhauser, eds. *Principals and agents: the structure of business*. Boston: Harvard Business School Press, p. 37-51.
- Blomqvist, A., 1991. The doctor as double agent: information asymmetry, health insurance, and medical care. *Journal of Health Economics*, 10(4), pp. 411-432.
- Bloor, M., Venters, G. & Sampier, M., 1977. Geographical variations in the incidence of operations on tonsils and adenoids. *Journal of Laryngology Otology*, 92(9), pp. 791-801 (part 1) and 92(10), pp. 883-895 (part 2).
- Bradshaw, B., 2013. *Cost of Governments NHS upheaval*. [Online] Available at: <http://www.benbradshaw.co.uk/cost-of-governments-nhs-upheaval/> [Accessed 24 January 2014].
- Bridgewater, B. et al., 2011. *Maintaining patients' trust: modern medical professionalism*. Henley-on-Thames: Dentrite Clinical Systems Limited.
- Campell, D., 1969. Reforms as experiments. *American Psychologist*, 24(4), pp. 409-429.
- CMS (Centers for Medicare & Medicaid Services), US Department of Health and Human Services, 2013. *Premier hospital quality incentive demonstration*. [Online] Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalPremier.html> [Accessed 7 January 2014].
- Cooper, R., 2013. *Inequality is at the core of high health care spending: a view from the OECD*. [Online] Available at: <http://healthaffairs.org/blog/2013/10/09/inequality-is-at-the-core-of-high-health-care-spending-a-view-from-the-oecd/> [Accessed 23 December 2013].
- Cooper, Z., Gibbons, S., Jones, S. & McGuire, A., 2011. Does hospital competition save lives? Evidence from the English NHS patient choice reform. *Economic Journal*, 121(554), pp. F228-F260.
- Codman, E., 1917. *A study of hospital efficiency: as demonstrated by the case report of the first five years of the private hospital*. [Online] Available at: <http://www.cms.gov/Medicare/Quality-Initiative-Patient-Assessment-Instruments/HospitalQualityInits/HospitalPremier.html> [Accessed 24 January 2014].
- Cutler, D. & Sheiner, L., 1999. The geography of Medicare. *American Economic Review*, 89(2), pp. 228-233.
- Deming, W., 2000. *The new economics for industry, government, education*. Second ed. Cambridge, MA: The MIT Press.
- DHSS (Department of Health and Social Security), 1976. *Priorities for health and personal social services in England: a consultative document*. London: The Stationery Office.
- DHSS, Great Britain Welsh Office, 1979. *Patients first: consultative paper on the structure and management of the National Health Service in England and Wales*. London: HMSO.
- Doran, T. et al., 2008. Exclusion of patients from pay-for-performance targets by English physicians. *New England Journal of Medicine*, 359(3), pp. 274-284.
- Dranove, D. & Satterthwaite, M., 1992. Monopolistic competition when price and quality are imperfectly observable. *RAND Journal of Economics*, 23(4), pp. 518-534.
- Edwards, N., 2010. *The triumph of hope over experience*. London: NHS Confederation.
- Fisher, E., 2003. Medical care—is more always better? *New England Journal of Medicine*, 349(17), pp. 1665-1667.
- Fleetcroft, R. & Cookson, R., 2006. Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Services Research and Policy*, 11(1), pp. 27-31.

## *Contracting for Quality in the NHS*

- Francis, R. (chairman), 2013. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London: The Stationery Office: .
- Gaynor, M., Moreno-Serra, R. & Propper, C., 2010. *Death by market power: reform, competition and patient outcomes in the National Health Service*. Working paper w16164, Cambridge, MA: National Bureau of Economic Research.
- Glover, J.A., 1938 The incidence of tonsillectomy in school children. *Proceedings of the Royal Society of Medicine*, 31(9), pp. 95–112.
- Harley, M. et al., 2005. Was Rodney Ledward a statistical outlier? Retrospective analysis using routine hospital data to identify gynaecologists' performance. *British Medical Journal*, 330, p. 929.
- Harris, J., 1977. The international organization of hospitals: some economic implications. *The Bell Journal of Economics*, 8(2), pp. 467-482.
- Jha, A., Joynt, K., Orav, E. & Epstein, A., 2012. The long-term effect of Premier pay for performance on patient outcomes. *New England Journal of Medicine*, 366(17), pp. 1606-1615.
- Kahneman, D. & Tversky, A., 1979. Prospect theory: an analysis of decision making under risk. *Econometrica*, 42(2), pp. 263-291.
- Kind, P. & Williams, A., 2004. Measuring success in health care—the time has come to do it properly. *Health Policy Matters*, No. 9.
- Kohn, L., Corrigan, J. & Donaldson, M. eds., 2000. *To err is human: building a safer health system*. Washington, DC: National Academy Press.
- Lunacy Act. (1845) 8 & 9 Vict., c. 100. London: House of Commons.
- Maynard, A. & Bloor, K., 2003. Trust and performance. *Journal of the Royal Society Medicine*, 96(11), pp. 532-539.
- Maynard, A., 2012. The powers and pitfalls of pay for performance. *Health Economics*, 21(1), pp. 3-12.
- McKinsey & Company, 2009. *Achieving world class productivity in the NHS 2009/10-2013/14: detailing the size of the opportunity*. London: Department of Health.
- Meacock, R., Kristensen, S. & Sutton, M., 2014. The cost effectiveness of using financial incentives to improve provider quality: a framework and an application. *Health Economics*, 23(1), pp. 1-13.
- Merrison, A. (chairman), 1979. *Royal Commission on the National Health Service: report*, London: The Stationery Office.
- Newhouse, J. & Garber, A., 2013. Geographical variations in health care spending in the United States: insights from the Institute of Medicine report. *Journal of the American Medical Association*, 310(12), pp. 1227-1228.
- Newhouse, J. et al. eds., 2013. *Variation in health care spending: target decision making, not geography*. Washington, DC: National Academies Press.
- Nightingale, F., 1863. *Notes on hospitals*. Third ed. London: Longman, Green, Longman, Roberts and Green.
- O'Neill, O., 2002. Question of trust. *BBC Reith Lectures*. [Online] Available at: <http://www.bbc.co.uk/programmes/p00ghvd8> [Accessed 23 December 2013].
- Percival, T., 1803. *Medical ethics: or code of institutes and precepts, adapted to the professional conduct of physicians and surgeons*. Manchester: S. Russell.
- Rettmeier, A. & Wang, Z., 2012. Regional variations in medical spending utilization: a longitudinal analysis of the US Medicare population. *Health Economics*, 21(2), pp. 67-82.
- Serumaga, B. et al., 2011. Effect of pay for performance on the management and outcomes of hypertension United Kingdom: interrupted time series study. *British Medical Journal*, 32, p. d108.
- Sheiner, L., 2013. Why the geographic variation in health care spending can't tell us much about the efficiency or quality of our health care system. *Finance and Economic Discussion Series*. Washington, DC: Federal Reserve Board.

*Contracting for Quality in the NHS*

Smith, A., 1759, *The theory of moral sentiments*, Raphael, D. & Macfie, A. eds. Oxford: Oxford University Press, 1976.

Sutton, M. et al., 2012. Reduced mortality with hospital pay for performance in England. *New England Journal of Medicine*, 367(19), pp. 1821-1828.

Syverson, C., 2011. What determines productivity? *Journal of Economic Literature*, 159(2), pp. 326-365.

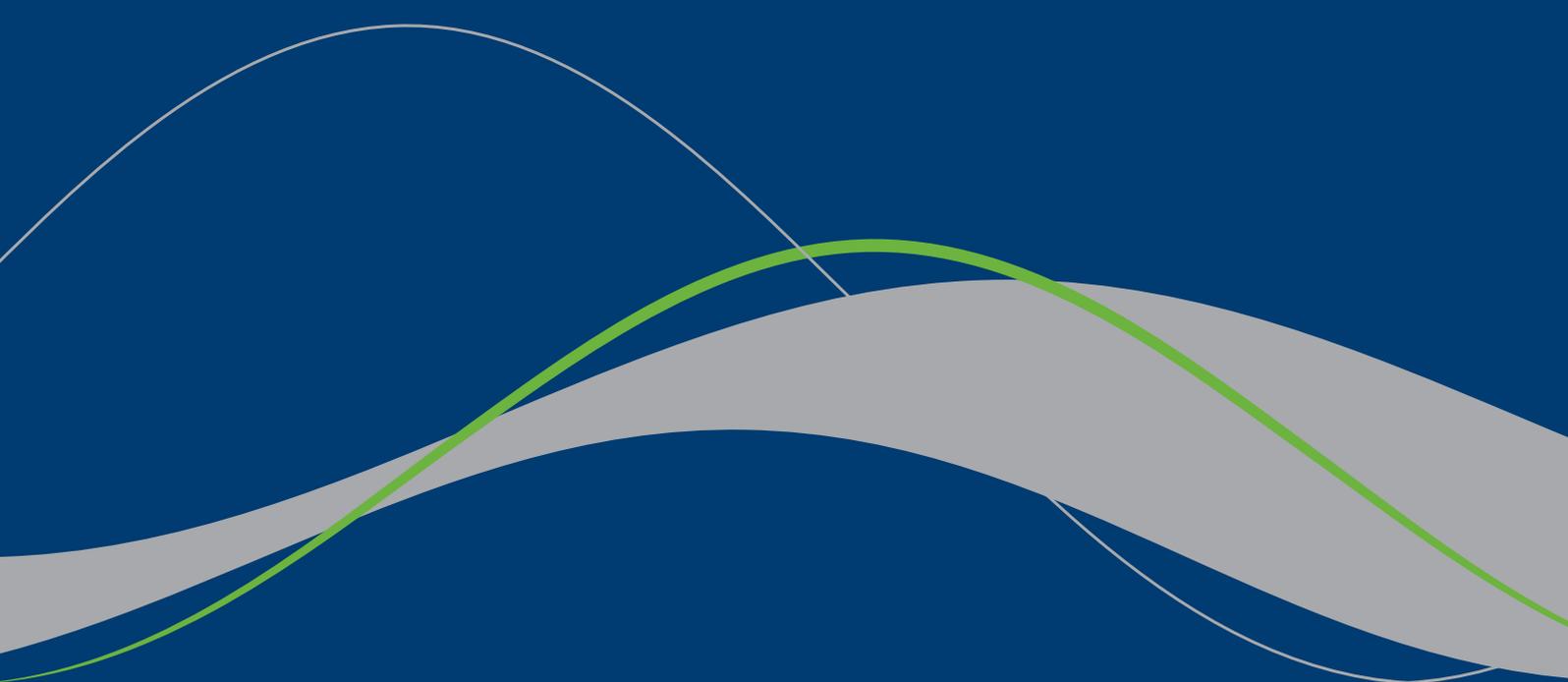
Timmins, N., 2012. *Never again? The story of the Health and Social Care Act 2012*. London: The Kings Fund.

Welch, H., Schwartz, L. & Woloshin, S., 2011. *Over-diagnosis: making people sick in the pursuit of health*. Boston: Beacon Press.

Wennberg, J., 2010. *Tracking medicine: a researcher's quest to understand health care*. New York: Oxford University Press.

Williams, A., 1978. Need—an economic exegesis. In: A. Culyer & K. Wright, eds. *Economic aspects of health services*. London: Martin Robertson, pp. 32-45.

Williams, A., 1988. Priority setting in public and private health care: a guide through the ideological jungle. *Journal of Health Economics*, 7(2), pp. 173-183.



Office of  
Health  
Economics

Research & Consulting

Office of Health Economics, Southside, 7th Floor, 105 Victoria Street, London SW1E 6QT, United Kingdom  
Tel: +44 (0)20 7747 8850 • Fax: +44 (0)20 7747 8851 • Email: [ohgeneral@ohe.org](mailto:ohgeneral@ohe.org) • Blog: [news.ohe.org](http://news.ohe.org) • Website: [www.ohe.org](http://www.ohe.org)