

viewpoint

Competition, integration
and incentives:
the quest for efficiency in
the English NHS

Alan M. Garber/Cambridge, Massachusetts

Contents

1	About the author
2	Acknowledgements
4	Foreword
6	Introduction
11	The Coalition Government's plan to repair the NHS
14	Barriers to efficiency in the NHS
19	The tentative embrace of market solutions in the NHS
23	Competition in health care markets
26	Integration for efficient care
31	Can competition and integration coexist?
34	Incentives for quality and efficiency
52	Effects of incentives in the English NHS
55	Health reform in England and the US
62	Concluding thoughts
66	References

Nuffield Trust Viewpoints provide a platform for UK and international health leaders to explore, discuss and debate critical health care reform issues. The views expressed are the authors' own, and do not necessarily represent those of the Nuffield Trust.

This Viewpoint forms part of the Nuffield Trust's work programmes on new forms of care, and on competition and market mechanisms.

As the pressure increases to improve efficiency and enhance quality, the NHS is looking towards new ways of organising and delivering care. We are examining the potential of new forms of care provision that are intended to benefit patients and taxpayers.

The Government health reforms have paved the way for a significant extension of market mechanisms in health care. Our experts are helping to build up the evidence base by examining where competition might help improve quality and efficiency.

Our award-winning website brings together our research and analysis in these areas at www.nuffieldtrust.org.uk/integrated-care and www.nuffieldtrust.org.uk/competition

About the author



Professor Alan Garber

Professor Alan Garber is the Provost of Harvard University, a post he assumed in September 2011. Before that he was Henry J Kaiser Jr Professor at Stanford University, where he was also a Professor of Medicine and a Professor of Economics and of Health Research and Policy; he also directed the Center for Health Policy and the Center for Primary Care and Outcomes Research. He was a staff physician at the VA Palo Alto Health Care System. He is a member of the American Society for Clinical Investigation, the Association of American Physicians and the Institute of Medicine of the National Academy of Sciences; he is also a Fellow of the Royal College of Physicians and of the American College of Physicians.

Alan served for 19 years as the founding Director of the Health Care Program of the National Bureau of Economic Research, where he remains a research associate. He was also the Chair of the Medicare Evidence Development and Coverage Advisory Committee (Centers for Medicare and Medicaid Services), and a member of the Panel of Health Advisers of the Congressional Budget Office, the National Advisory Council on Aging (National Institutes of Health), and many committees of the National Institutes of Health and of the National Academies.

His work addresses methods for improving health care delivery and financing, particularly for the elderly. It encompasses technology evaluation, analysis of the causes of health expenditure growth and health care productivity. A *summa cum laude* graduate of Harvard College, he received his PhD in economics from Harvard and an MD with research honours from Stanford, and trained in medicine at Brigham and Women's Hospital.

The Nuffield Trust designated Professor Garber as the 2011 Rock Carling Fellow. Part of his fellowship has included the writing of this Viewpoint.

Acknowledgements

This discussion of the challenges facing the English NHS grew from the extended visit my family and I made to the UK during the summer of 2009, under the sponsorship, financial support and kind hospitality of Jennifer Dixon and her colleagues at the Nuffield Trust. Jennifer arranged for me to meet many of the leading lights of health policy in England. Her insights into the past, present and future of health policy in the UK were both penetrating and prescient, and I am grateful for her encouragement and support throughout this project.

The ideas expressed here were informed by conversations with past or current NHS officials and other experts in Great Britain and in the United States. I am grateful to all of them for both their candour and their many insights. They include Martin Bardsley, Nicola Bent, Dame Carol Black, Kalipso Chalkidou, Francis (Jay) Crosson, Penny Dash, Andrew Dillon, Tim Doran, Jean Drouin, Nigel Edwards, Chris Ham, Nicolaus Henke, Ron Kerr, Richard Lewis, Carol Longson, Ruth McDonald, Andrew McKeon, Lucy Moore, James Mountford, Bill Moyes, Heather O'Meara, Keith Palmer, Carol Propper, Sir Michael Rawlins, Rebecca Rosen, Judith Smith, Peter Smith and Barbara Young.

Nomita Divi and Christine Geibel assisted with many facets of the manuscript preparation. I also benefited greatly from conversations with Jay Bhattacharya, Anita Charlesworth, Martin Conner, Jennifer Dixon, Alain Enthoven and Douglas Owens. Each of them made insightful comments on earlier drafts of the manuscript, leading not only to a clearer exposition but a more accurate depiction of a constantly evolving set of government policies in both England and the United States (US).

My research in closely related areas, which informed the views expressed here, was supported by the US National Institute on Aging, the Agency for Healthcare Research and Quality, and by an Investigator Award in Health Policy Research from the Robert Wood Johnson Foundation. This work was conducted while I was a staff physician at the US Department of Veterans Affairs and a faculty member at Stanford University. I am grateful to both institutions for

providing a truly extraordinary intellectual environment and set of colleagues. It is hard to imagine a better place to have built a career that combines medicine, economics and health policy. As I leave these institutions after 25 very happy years, I am keenly aware of the debt that I owe to my mentors, colleagues, students and everyone else who made the Center for Health Policy and Center for Primary Care and Outcomes Research at Stanford so successful.

Special thanks to my wife, Anne, who made our visit to England – and so much else – possible.

Foreword

The Government's NHS reforms pave the way for more competition and a more locally managed health service. They also take place at a time when the NHS in England is facing what is often dubbed the 'Nicholson challenge' – that of saving an estimated £15 to £20 billion by 2015. Achieving savings of this level will require the NHS to move from a notion of technical efficiency towards allocative efficiency beyond the walls of providers, and a more radical overhaul of how services are designed and delivered. Critical to achieving this is creating the right incentives for the NHS to develop, in ways that promote creativity and innovation.

Alan Garber, currently of Harvard University and formerly a Professor of Medicine at Stanford, is the Nuffield Trust's current Rock Carling Fellow. He brings an economist's view to the debate, together with considerable experience and understanding of the US health system. Avoiding easy judgments his research, presented here, explores the question of how to find the right balance of competition, integration and incentives for the NHS to promote innovation in service delivery. He considers this within some future parameters, including health services that will be increasingly locally shaped; expenditures pegged at sustainable levels of growth; and huge pressures to extract as much quality as possible for every pound spent.

This Nuffield Trust Viewpoint reflects the Trust's aim of bringing international expertise to discussion of NHS reforms. I do hope you will find it a useful contribution to that discussion. To find out more, visit www.nuffieldtrust.org.uk/international-comparisons.

Dr Jennifer Dixon, Director, Nuffield Trust, December 2011

This Nuffield Trust
Viewpoint reflects the Trust's
aim of bringing international
expertise to the NHS
reform debate

Introduction

“I doubt the NHS will be able to deliver on all these demands and expectations. At least on the present lines of fiscal policy, there won’t be enough money. There seems to be little support for higher taxes and no easy source of funds elsewhere in the budget to raise health services spending to the level of other northern European countries.”

(Enthoven, 1999)

The clash between the desire to improve services and the need to confront straitened financial circumstances is a central policy problem – perhaps the central problem – for the UK’s Coalition Government. But these words, written nearly 12 years ago by Alain Enthoven for his Nuffield Trust Rock Carling lecture, refer to a time when the NHS was under a different government and the world economy was not mired in financial uncertainty. Despite the many changes that the NHS has experienced during the last decade, the Coalition Government confronts a set of challenges strikingly similar to those that my colleague so ably analysed: these recurring challenges have been addressed, but never fully overcome.

During the past decade, the English NHS largely accomplished the Labour Government’s goal of reducing queues and raising the standard of care more generally (Bevan and Hood, 2006). Yet the UK, like all the wealthy nations of the world, faces the need to improve the efficiency of its system of health care. In a 16 May 2011 speech addressing delays in implementation of the Coalition Government’s proposed health reforms, Prime Minister David Cameron observed that “the NHS is the most important thing to Britain’s families,” adding, “It’s the most important thing to my family too.” But the first item on his list of the problems faced by the NHS was “waste and inefficiency” (Cameron, 2011). What a government leader describes as ‘waste’, of course, can easily be the third-line, unproven cancer treatment that a patient and their physician believe offers the only hope against a life-threatening disease. As the Coalition Government has come to understand, it is difficult to eliminate waste and inefficiency while claiming to protect every aspect of the health system that the public so passionately defends.

The quest for efficiency is occasionally dismissed as an ill-disguised excuse for engaging in unpopular cost-cutting measures. In the US,

such rhetoric has had a profound effect on the political discourse. Yet such a representation is neither accurate nor fair. Cutting costs is not easy, but it poses fewer technical difficulties than improving efficiency. At a mechanistic level, cost reduction is straightforward, especially in a health system with central budgets: to control costs, simply restrict the available funds. At a political level, however, public reaction impedes and can put a complete stop to such efforts. The expenditures, even if labelled as 'waste', go to a set of services that some people value and which they expect will always be available. Furthermore, the public expects the quality of services not only to be maintained, but to improve steadily with time. As Cameron's speech suggests, political resistance will overpower any attempted reforms that are not perceived to maintain or improve quality.

Indeed, concerns about the reaction to cost reduction were so powerful that the Coalition White Paper (Department of Health; DH, 2010) promised to boost real spending for health care at a time when austerity was the order of the day for other parts of the budget. The Coalition Government, according to the document, aimed to 'reduce mortality and morbidity, increase safety, and improve patient experience and outcomes for all' (DH, 2010: p4). At the same time, the White Paper included the claim that 'the NHS will need to achieve unprecedented efficiency gains... to meet the current financial challenge and the future costs of demographic and technological change' (p5). Even though specific recommendations from the White Paper have already been delayed, softened or abandoned, efficiency gains are the solution the government will seek in order to bring costs under control while maintaining the quality of care. Nothing else is likely to succeed in the political arena. Thus the central health policy issue in the UK can be stated simply: How might the NHS improve overall efficiency, at a time of growing financial constraint?

In the discussion that follows, I approach the quest for efficiency from a basic economic perspective that I hope does not veer too far into the abstract. Simple theories, economic or otherwise, that apply well in other settings often convey little insight and earn their authors no credibility when applied to health care. Although it would be a mistake to rely too heavily on such theories, in health care we often commit the opposite error: our views are often so firmly grounded in an informal empiricism that we ignore some of the insights that can come from a perspective informed by theory and observations from other domains.

“The incentives that most limit what an organisation can do are those that are imposed from without.”

My comments reflect an economist’s perspective on health care, and that is why I focus on structures for health care delivery, the markets in which those structures interact and the incentives that influence individual actors. Incentives are the key to performance, but designing incentives is much harder than is commonly understood. As the financial crisis made abundantly clear, all too often explicit incentives are put in place without adequate appreciation of their ramifications. The inability to ensure that those most directly involved in crucial financial decisions – such as traders and executives of banks and investment firms – faced rewards and penalties that would cause them to act in the best interests of the firms’ owners (shareholders) was widely viewed as a fundamental cause of the crisis. In a sector in which financial incentives are not only of first-order importance, but other incentives such as social approval are often dismissed as having little importance, the alignment of the interests of owners and management is less complicated than in health care. Nonetheless, great minds have laboured without definitive success to resolve incentive problems (known also as principal–agent problems) in the financial sector. If incentive design is a challenge in the financial world, it is yet more daunting in the world of health care.

The incentives that most limit what an organisation can do are those that are imposed from without – the external rules under which organisations operate. They are important in the policy sense, since external incentives – when misdirected – make it impossible for organisations and individuals to adopt desired behaviour without suffering penalties. This is readily apparent in the US, where much health care is reimbursed on a fee-for-service basis. A hospital or physician delivering efficient care in a fee-for-service setting will often be punished financially because the external incentives reward volume rather than outcomes or value. Recognition of the adverse consequences of the payment mechanisms used by much of the federal Medicare programme and by many private health insurers was one of the principal motivations for the Affordable Care Act, the health reform bill that became law in 2010.

Internal incentives also matter greatly, and they are subject to the control of the organisation. Any organisation delivering health care must ensure that the members of the organisation – executives, partners, employees and everyone else who works on its behalf – perform in a manner that advances the organisation’s goals. This is one example of the principal–agent problem – how the principals, or

“England is hardly unique in restricting the ways that hospitals and physicians can organise care.”

owners, can structure agreements so that the corporate executives and other agents to whom they delegate much of the work of the organisation will advance the owners’ goals, even though the agents have much more detailed knowledge about the organisation and the circumstances it finds itself in, and even though their interests coincide imperfectly. Large, hierarchical organisations that have uniform internal incentives often can ensure greater transparency and knowledge of the incentives, and may be able to administer the incentives more easily. But they give up flexibility and adaptability; important strengths when the external environment changes rapidly. The monetary and non-monetary incentives used by an organisation, whether they are explicit or not, and whether the organisation is large and hierarchical or small and informal, are often the most important determinants of an organisation’s ‘culture’.

Organisational structure and culture are often viewed as the key to performance, but incentives and organisational behaviour are tightly linked. The discussion that follows touches on organisation, which is a focus of health reforms in England and, I believe, has received disproportionate emphasis. Organisational governance and form matter, but attempts to ‘fix’ organisation without creating appropriate incentives are unlikely to be successful. Furthermore, policies often directly or indirectly favour particular organisational forms, despite a lack of compelling evidence that they will lead to superior performance. Such policies can impede the development of better alternatives. For example, some countries discourage or explicitly prohibit corporate organisation, limiting the ability of hospitals and other health care organisations to raise capital. In the US, various laws restrict how physicians can organise, the types of contracts that physician organisations can sign, and the creation of entities that combine care delivery and insurance or commissioning functions.

Thus England is hardly unique in restricting the ways that hospitals and physicians can organise care. But government control comes at a cost: in markets that are regulated less stringently, firms are free to evolve towards forms that are efficient and therefore successful in a competitive environment.

Fully market-determined organisation of care may be a chimera, since no major health system is part of an entirely unregulated market. However, competitive elements are not only tolerated, but even promoted in centrally administered systems – thus the ‘internal

market' of the 1991 NHS reforms. Flexible organisation is an important aspect of competition. A flexible policy need not imply ignorance of the problems that can occur with unregulated – or lightly regulated – approaches to the organisation of care delivery. For example, integration of services typically requires both scope and scale – multiple specialties and services, and large size – which in turn can lead to market power and the attendant potential for anti-competitive behaviour. One approach is to prohibit the formation of integrated organisations or to strictly limit their size. But the need to balance the economies that can occur with large scale and scope against the threats to competition that occur with monopoly is well recognised, and a large body of literature as well as legal precedent offer approaches to managing such problems.

The quest for performance – and the need to consider incentives, competition and integration – motivate these comments. These topics are central to debates about the future of the English NHS. Although I touch on the current reform, my purpose is less to evaluate the changes overall than to view them from the perspective of incentives and the ways that they address the central challenges of health care delivery. Indeed, because I cannot claim a deep and intimate familiarity with the NHS, my observations about the details of the reform would lack the necessary nuance. Furthermore, at a time when there is so much uncertainty about the future of reform, conclusive statements about its prospects of success would be premature. And even if there were no uncertainty about the broad plan outlined in the White Paper and the changes outlined in the government's response to the NHS listening exercise (DH, 2011), the details of implementation are likely to be more consequential than the broad policy. Would the newly renamed clinical commissioning groups, for example, be able not only to manage themselves internally but also to strike agreements with trusts to ensure that they work together to increase the efficiency of care? Much will depend on the ability of the groups to form proper risk-sharing agreements and to exercise their purchasing power appropriately. That is likely to become clearer with further development of the reforms.

It is my hope that these comments, from the perspective of an American economist and physician who has followed developments in the UK with great interest and sympathy, will provide a useful framework for thinking about the broad rationale for key aspects of the reforms and also about the ways that they might be implemented.

The Coalition Government's plan to repair the NHS

The popularity of the NHS among the British public is real and deep. The British Social Attitudes survey (NatCen, 2010) reports that 64 per cent of the British public are either very or quite satisfied with the NHS – the highest level of satisfaction since the survey began, and part of a continuous upward trend since 2002 (Appleby, 2011a). Yet the NHS is not immune to the threats posed by government budget deficits and the uncertain future of the economy. In the lead-up to the May 2010 election, when nearly every category of public expenditure was subject to the prospect of deep cuts, no major candidate endorsed significant reductions in the NHS budget (Hutton and Penny, 2010; Chantrill, 2011). Taking health care off the table was remarkable because the NHS comprises the largest category of government spending in the UK, and its share of gross domestic product has risen accordingly. Other areas that important constituencies also favour, such as education, have experienced lower growth, especially as a share of the government budget, as can be seen in Figure 1, on page 12.

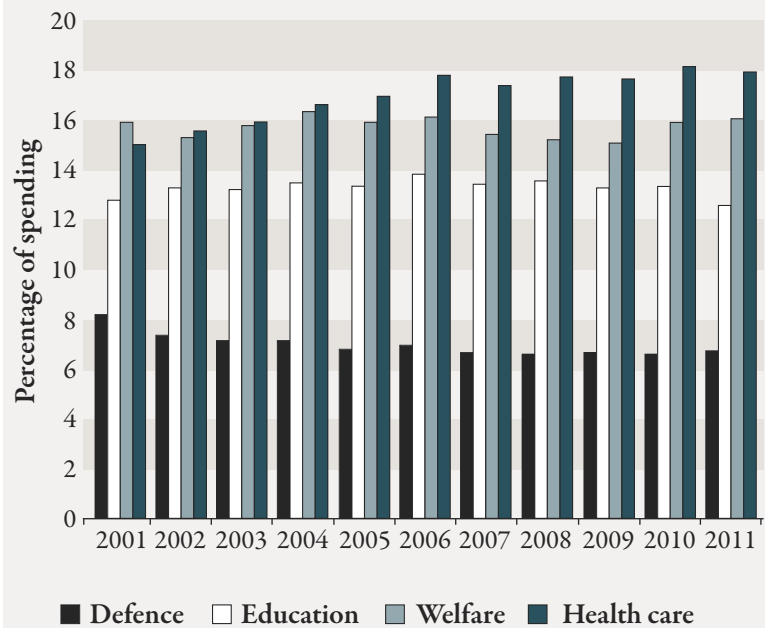
64%

of the British public are either very or quite satisfied with the NHS

It is uncertain how long public support will shield even this popular programme from the pressures of record government budget deficits. Rumbles of dissatisfaction with shrinking public services and education budgets may grow louder; challenging the privileged position that the NHS has held in the public eye. The performance of the NHS has become a central concern – if not *the* central concern – of UK policy.

Soon after the election, the Coalition Government described its vision for the NHS in a White Paper (DH, 2010). The recommendations it contained were hotly debated in Parliament and among the public: the phase-out of primary care trusts (PCTs) and strategic health authorities (SHAs), and the shift of decision-making authority to newly established consortia of general practitioners (GPs) raised questions of both principle and of practicality. Would clinical commissioning groups be as responsive to community needs? Would they have the general management expertise needed to run a complex organisation? Would they be capable of managing relations with hospitals, and would they coordinate effectively with specialists?

Figure 1: UK Government spending shares dedicated to health and other areas, 2001/11



Source: Chantrill, 2011

(Ham, 2010). Although the White Paper did not lay out detailed plans for implementing the policy it outlined, the fundamental premise was clear: greater responsiveness to the needs of individual patients, predicated on a belief that a greater degree of decentralisation and competition within a system of oversight, can lead to more efficient and higher quality care. These goals not only echo some aspects of previous Conservative governments, but also Tony Blair's efforts to introduce competition and choice in the NHS.

As an aspiration, improving NHS performance has broad appeal. Nevertheless, the White Paper was met immediately with questions, concern and scepticism. In contrast to the US, in the UK major health reforms have been put in place at least three times in the last 20 years, along with a number of more narrowly targeted initiatives. Even a seemingly minor change can disrupt a clinician's practice and necessitate costly adjustments by hospitals. Thus reforms have left

providers jaded and weary, and the public sceptical. And despite the Coalition Government's claims to the contrary, the shift it proposed to GP-based commissioning looked very much like the GP fundholding introduced by the last Conservative government – a sibling, if not an identical twin. But under the previous Conservative policy, only a minority of GPs became fundholders (Maddox, 1999), while the White Paper envisioned that '... every GP practice will be a member of a consortium...'. So, in this respect, the planned reforms were more sweeping. And, although the plans have been modified since the White Paper was released – including changing the name of the key commissioning entities from 'GP consortia' to 'clinical commissioning groups', with representation of specialists and other groups – the challenges remain the same. What steps will make the reform strategy successful?

Barriers to efficiency in the NHS

Notwithstanding occasional claims that the NHS practises rationing, limits patient choice and puts the needs of the nation ahead of the needs of its citizens, even in the US there is a great deal of respect for the (relative) efficiency of the NHS and its universal access to medical care. For many years, complaints about the NHS have been provoked in large part by the inability of improvements in NHS services to keep up with the expectations that accompanied rising incomes. As Simon Stevens, at the time the health policy adviser to Prime Minister Tony Blair, succinctly put it in 2004: 'since 1948 Britain's single-payer National Health Service (NHS) has given it universal coverage and overly effective cost containment' (Stevens, 2004). Long waits for surgery and other procedures became a growing and visible source of public dissatisfaction in the 1990s. As Stevens noted, this contributed to a 'growing tendency of the British media to substitute its long-standing stereotype of the NHS ("good") versus the US health system ("bad"), with an equally polemical comparison of the NHS ("bad") with continental Europe ("good")'.

In the 1990s, the NHS was indeed viewed as a low-cost, highly effective health system. But even then there were concerns about its efficiency. In a report conducted by the McKinsey Global Institute in the 1990s, a condition-specific comparison of health system efficiency showed that for some conditions the US, and at times Germany, performed better than the UK. That the US came off well seemed surprising. However, in this instance the apparent inefficiency of US health care was due more to high prices of input factors than to inefficient deployment of those resources (Baily and Garber, 1997). British patients with lung cancer, for example, were less likely to receive a staging CT scan than German and US patients. The more parsimonious approach to imaging might have led to erroneous 'down-staging' in which patients were only found to have disease too extensive to be treated surgically when they were already on the operating table. This finding suggested that constraints on the availability of CT scans – there were far fewer scanners per capita in the UK than in Germany or the US – led British patients to receive inappropriate operations more frequently than patients in other countries.

“With its predominantly fee-for-service payment system, medical care in the US tended to focus on reimbursable services.”

In other situations, however, the NHS seemed to offer superior performance. Despite claims of critics that poor integration of care leads to inefficiency in the NHS, that was not the case for diabetes care in the McKinsey study. Here the NHS appeared to produce better outcomes than the US health system at lower cost through an effective integration strategy. Providers in the UK simply seemed to manage diabetes more efficiently than providers in the US. Better performance may well have resulted from better coordination of care for diabetics in the UK, where GPs referred patients with more challenging disease to multidisciplinary clinics. With its predominantly fee-for-service payment system, medical care in the US tended to focus on reimbursable services, such as procedures, laboratory tests and physician office visits. Unsurprisingly, physicians had little interest in delivering services that were not eligible for payment. Thus, for most of the health care system, multidisciplinary team care and the services of, for example, nurse-educators and dieticians, had a limited role. In the UK, by contrast, diabetes management had been identified as a target of a national initiative and special funding had been allocated for such programmes (Baily and Garber, 1997); a US–British comparison showed better outcomes for diabetic patients in the UK several years later (Banks and others, 2006).

To the extent that under-funding limited the performance of the NHS, the Blair Government took decisive steps to address the problem. It dramatically increased public investment in the NHS, increasing per household funding levels from £3,422 when it took office, to £4,255 by 2008 (Thompson, 2009). The investments had measurable impacts. The median waiting time for inpatients, for example, fell from about 17 weeks in June 2007, to about eight weeks less than a year later (Appleby, 2011b).

The Quality and Outcomes Framework (QOF), which linked GP payments to performance measures, along with other initiatives, helped to improve ratings on quality measures and to cut waiting times for procedures and specialist consultations. The QOF and other programmes also improved access to out-of-hours care, and criticism of inadequate access to capital for expenditures on scanners and other high-cost technologies became more muted. But these programmes were costly, leading to questions being asked about whether the same improvements might not have been obtained at much lower cost. And, more generally, there were doubts about the

“PCTs have few instruments to ensure that specialists and hospitals are used effectively and efficiently.”

coordination of care and the ability of PCTs to manage commissioning effectively.

The modern NHS was never known for highly integrated care, in part because hospital-based specialists and primary care physicians worked in different institutions and with different modes of compensation. The lack of integration, in which one physician may have reasons to minimise the use of hospital services, while another has incentives to increase their use, can make it difficult to achieve efficiency (Fox, 2002). A widely-read study created controversy on both sides of the Atlantic when it suggested that Kaiser Permanente, the American health maintenance organisation, provided high-quality medical care more efficiently than the NHS (Feachem and others, 2002). The superior performance was attributed in part to tighter care integration.

Fragmentation in the NHS would seem paradoxical to anyone who thinks of it simply as a unitary governmental health system, in which the government both pays for health care and, to a great extent, provides it. The principal requirements for integration seem to be inherent in any such system. Organisational structure and incentives could be designed to promote coordination of all aspects of service, and it should be possible to invest in the necessary physical and information infrastructure, as well as personnel, to make it work. But although health care providers in the NHS are subject to oversight and regulation, they exercise considerable independence. In principle, the central body organising care before the recent reforms was the PCT, which plays the role of commissioner or insurer. Unlike an insurer in the US or in many other nations, however, a PCT is responsible for funding the care of all the people living within a defined area. A PCT both has a pure monopoly over primary health care supply and is a monopsony (sole) purchaser of local primary care services.

But although the PCT administers contracts with GPs for personal medical services, it typically has little control over GP decisions or the expenditures that result from specialty referrals or inpatient care. When GPs refer their patients to hospital-based specialists, or their patients have unscheduled admissions to hospital, the PCT bears the cost, which typically account for about 50 per cent of their overall spending. Despite the magnitude of these expenditures, PCTs have few instruments to ensure that specialists and hospitals are used

“Distinct and often incompatible incentives among the key players conflict directly with efforts to integrate care.”

effectively and efficiently. Furthermore, many of these secondary services are reimbursed according to ‘Payment by Results’, a fixed-fee schedule or tariff. Specialists and GPs lack consistent incentives to limit utilisation of services they are not at financial risk for.

PCTs have also been responsible for public health measures. With such diverse missions, and a limited set of levers to influence provider behaviour, their effectiveness as commissioners has frequently been questioned. A PCT cannot easily reallocate resources to programmes that contribute more to the health of the community it serves, in large part because it has so little control over the activities that drive provider budgets. Thus, even if a PCT could readily identify the optimal allocation of health resources, it would not necessarily have the authority and instruments to implement it.

Distinct and often incompatible incentives among the key players – GPs, PCTs and hospitals with their specialists – conflict directly with efforts to integrate care. In theory, at least, PCTs might be able to restructure incentives to ensure tighter integration between GPs, specialists and hospitals. But PCT directors have diverse views of their abilities to negotiate contracts that would overcome some of the incompatibilities in incentives. For example, some have tried to redefine general practice and to facilitate closer interaction between specialists and GPs, in a structure akin to a multi-specialty group practice in the US (Burke, 2010). Others have felt that such attempts at integration would not be encouraged, or even tolerated, by the Department of Health.

If integration is elusive for many PCTs and locations, there are hopeful exceptions. Ham and Smith have described several promising efforts to integrate care in England, in such diverse PCTs as Nottingham, Torbay, Trafford and Redbridge (Ham and Smith, 2010). The Redbridge vision for care integration encompasses virtually all health and community services, based upon ‘polysystems’ whose constituents include NHS Redbridge as well as ‘GP practices, health provider organisations, pharmacists, hospital care clinicians and trusts, voluntary sector, independent sector and other providers’ (NHS Redbridge, 2011). In both Trafford and Redbridge, a prime motivation for the movement towards integrated care was the recognition that expenditures for acute (hospital-based) services would need to be controlled better. In each of these instances, integrated information technology (IT) and redefined, better aligned

financial relationships are central to the integration effort. This is one reason why there is disagreement about whether such arrangements are generally permissible within the NHS. Some have argued that misaligned incentives are intrinsic to Payment by Results and cannot be overcome without fundamentally changing its structure (Ham and Smith, 2010). But if initiatives to organise and coordinate care succeed despite such obstacles, it suggests that there could be even greater success if providers and commissioners were allowed greater freedom to organise or reach agreements that better align incentives.

Will a shift of care towards new commissioning groups facilitate integration? That depends heavily on the ability of commissioners and other entities to formulate contracting arrangements that support common interests and goals. A critical feature will be the ability to share savings from efficiencies that integration might deliver.

The tentative embrace of market solutions in the NHS

Even though the NHS is a single governmental entity, the activities it finances are not necessarily integrated. GPs, hospital trusts and PCTs have faced distinct incentives. Efforts to encourage different care providers to coordinate services have been neither widespread nor sustained. Indeed, there has been little consensus about the payment mechanisms that would promote integration. In some quarters, this undoubtedly reflects ambivalence towards, if not outright opposition to, the use of financial incentives to drive behaviour.

Whatever the cause, the lack of coordination makes the NHS less integrated, and presumably less efficient, than it would otherwise be. How might the NHS adopt more integrated approaches to care, and to what extent is care integration consistent with the benefits that exchange in markets is expected to bring?

Here again my colleague Alain Enthoven has offered insights into the design of health care organisations and markets, and how efficiency might be promoted. Known as the architect and theorist of managed competition, he has promoted market-based approaches to health care delivery in the US, the UK and in other nations (Enthoven, 1978a, 1978b, 1993; Enthoven and van de Ven, 2007). His work had great influence on market-based reforms in both Conservative and Labour governments. He has consistently sought to introduce competitive elements to promote efficiency while maintaining elements of regulation to address market failure, particularly adverse selection. Thus his pro-competition, pro-market stance is heavily qualified.

In his Rock Carling lecture, Enthoven was critical of the regulatory oversight of internal markets in the NHS; raising doubts about the handling of monopoly power. He noted that requiring prices to equal costs simply gave providers with monopoly power the incentives to allow their costs to rise, ‘putting money into prestige-enhancing technology, pay rises, improved working conditions’ (Enthoven, 1999). His criticism extended to regulation as well; he noted that many of the regulations imposed during the 1990s fundamentally weakened competition. For example, trusts were not allowed to generate and save surpluses to ‘finance a competitive challenge to another trust’, thereby prohibiting one of the mechanisms that

successful organisations might use to increase their scale and serve more patients.

Thus a reluctance, perhaps not always conscious, to promote unfettered competition characterised regulatory policy even under a Conservative government. Some perceived the risks of competition to be great enough to justify regulations that would sacrifice some of the benefits. And even in the heyday of internal markets, the NHS remained a centralised organisation with considerable control over the arrangements made at the local level.

Of course, there are fundamental questions about what competition can be expected to achieve if key requirements of competition are absent. Prices under perfect competition are set by markets, not by decree. An increase in supply or a decrease in demand ordinarily lowers prices, while administered prices, such as the tariffs under Payment by Results, seldom respond quickly nor do they necessarily reach the same results. Advocates of administered prices do not necessarily view this as a flaw, since their goal is not to duplicate market outcomes. Administered prices may reflect political pressures, bureaucratic considerations, or even policy goals such as a desire to subsidise innovation, rather than a proper accounting of supply and demand. Another characteristic of competition is free entry and exit, which means that any provider – public or private – should be able to offer services and compete on equal terms. Although there is a private market for health services in England – and private providers have a role in the NHS – it remains limited. These are among the reasons why a pro-competitive policy in the context of the NHS or another highly regulated environment does not necessarily lead to the prices, quantity and quality of care that would be observed in a market that contained more elements of competition.

It is fair to ask what competition, or a pro-competitive policy, might do for the NHS. The pro-competition view is motivated by the belief that competition can bring results (Stevens, 2011).* Many have

* Defining competition, even in theory, has many nuances. Among economists, the meaning of ‘perfect competition’ has evolved over time (Stigler, 1957; McNulty, 1967). A common approach today builds upon the notion that competition means that a seller of a product or service faces great, if not perfect, elasticity of demand. In other words, under perfect competition, a seller cannot raise prices by very much without facing a sharp drop in demand. This does not always correlate perfectly with the number of sellers in a market.

“Studies suggest that competition leads to improvement in at least some measures of quality.”

argued that competition, even in a heavily regulated market, can lead to better quality at lower cost. Although general beliefs about the desirability of competition, rather than specific studies, undoubtedly have the greatest influence on opinions about its suitability in health care, a body of research has attempted to measure its effects. Studies of the impact of competition on health outcomes have reached mixed conclusions (Shortell and Hughes, 1988; Propper and others, 2003; Propper and others, 2004; Gaynor and others, 2010). Inconsistent findings across studies result in part from the varying circumstances they consider. In some studies, prices are fixed, while in others they are not; we would not expect similar behaviour in the two different conditions. In addition, researchers do not always use the same definitions of the market or the same measures of competition, nor do they always focus on the same disease conditions or outcome measures.

One of the greatest limitations of studies on the effects of competition is the use of research designs that make it difficult to distinguish cause and effect. The same forces that give rise to variation in the number of hospitals (usually the object of study) or physicians in a geographically defined market might also lead to variation in health outcomes – even if competition itself played no direct role in determining survival and other markers of health benefit. Studies using the most sophisticated designs, however, such as Kessler and McClellan (2000), Bloom and others (2010), and Gaynor and others (2010) suggest that competition leads to improvement in at least some measures of quality. A study of competition among GPs in the NHS reached similarly positive findings (Croxson and others, 2001).

The measures of competition used in empirical studies are constrained by the available data. The best of them tend to be measures of the number of competitors or standard market concentration measures. A more relevant gauge of competition, or the lack thereof, is also difficult to measure: the ability to raise prices without suffering a significant erosion of demand. If a producer faces inelastic demand, or demand that changes little in response to price changes, it is able to engage in monopolistic behaviour.

Empirical measures of market concentration are undoubtedly correlated with the ability to raise prices above competitive levels, but from the very start they must cope with the difficult issue of defining the relevant market. In the US, the presence of

multiple hospitals in an area is often sufficient grounds to suggest that conditions are consistent with vigorous competition. But if one hospital is considered significantly more attractive for any reason – for example, the surgeons who operate there have an outstanding reputation – its competition may not be a local hospital but rather other prestigious hospitals hundreds of miles away. Under such conditions, standard measures of market concentration are misleading.

Competition in health care markets

Most health policy discussions are not predicated on the belief that health care markets conform to the competitive ideal. They begin with the assumption that in medical care genuinely free competition does not and could not exist, if only because informational failures and the asymmetry between the knowledge of doctor and patient deny basic requirements for market efficiency (Arrow, 1963). The desire for protection from the financial consequences of illness and injury makes health insurance ubiquitous, which in turn means that prices (of insured items) are subsidised and patients tend to over-consume care. In every nation, markets for health care are heavily regulated and/or the government plays a major role as an insurer or provider of health services. So whatever its virtue as a normative ideal, perfect competition does not characterise the delivery of health care in any wealthy nation, including the US.

Whether a market can be said to differ greatly from perfect competition – in other words, to be characterised by monopoly or monopolistic behaviour – also depends heavily on the definition of the product or service that is sold in the market. In a sense, when geography and quality are considered product characteristics, the example of the higher-quality hospital is a special case of product differentiation. Low-quality and high-quality hospitals may provide services that consumers see as fundamentally different, so they do not compete directly. And more generally, the more differentiated the product, the less likely that a market will exhibit the behaviours that lead to direct and severe price competition. Thus the profound variation in medical products and services that are transacted has many consequences, often leading to reduced competition.

For example, mattresses are broadly similar products, and there are several mattress manufacturers, so there should be aggressive price competition. Yet mattress manufacturers have devised strategies to diminish price competition among very similar products. Many mattress sellers have price guarantees, offering to refund more than the difference in purchase prices if the customer finds the same mattress at another seller for a lower price. These refunds are seldom

“Highly aggregated contracting is often more efficient than disaggregated, customised forms of care.”

collected because mattress manufacturers assign different model numbers to nearly identical mattresses. This strategy discourages shopping for the lowest price because it is costly to learn whether the differences between two mattresses sold at different retailers are meaningful.

By diminishing the salience of price comparisons, pricing obscurity reduces price competition between retailers, which in turn raises the prices that manufacturers can charge. Mattress markets, in other words, are characterised by differentiation based on the model number, in addition to whatever physical differences exist among the products. Thus a relatively homogeneous product line becomes one with many differentiated products that serve as imperfect substitutes, even when the differences among products are irrelevant to the consumer.

When consumers are unable to distinguish between similar products or similar services, often competition over more disaggregated products and services will not only lead to inferior choices, but also increase the costs of production. Consider purchasing a house or carrying out a major renovation. The home buyer could engage a general contractor who would have full responsibility for selecting subcontractors and all of the materials used to build the house, subject to a general agreement about the types of materials (for example, electrical outlets, lighting and plumbing fixtures, walls, paint, floors) to be used and the overall quality of construction. To the extent that the quality of the workmanship and materials is hard to observe, the incentives of the contractor and the purchaser will deviate. With a fixed-price contract, the cost-minimising contractor will tend to under-provide quality but also to produce at low cost.

A knowledgeable home buyer might prefer to choose every component of the house and to contract with electricians, plumbers and carpenters directly. Such arrangements give the purchaser greater control over each detail of the house and its quality, but unless the purchaser has extraordinary expertise, a contractor will likely be more efficient, not only by making better choices of parts and labour, but by obtaining them at lower cost.

Similarly, highly aggregated contracting – for example, contracting with an integrated health care organisation to provide comprehensive care services for a fixed rate – is often more efficient than the

disaggregated, customised forms of care possible when the service being transacted is a single procedure or product such as a drug. The most important reason for aggregation is the buyer's lack of knowledge of the need for, or quality of, different components of care. But making choices at the individual service level will often fail for simpler reasons: consumers seldom know the cost consequences of choices when markets are defined by individual services.

The total cost of having nearly any surgical procedure, for example, would include expenditures for laboratory tests, imaging studies, hospital days, supplies, drugs and the services of ancillary personnel such as respiratory therapists. A patient shopping for a knee replacement would need to know the expected expenditures for all of these components of care in order to infer the total cost of a procedure – if the services and products were purchased at the disaggregated level. Patients in England do not pay for care directly, of course, yet PCTs do, and sellers paid by the unit of service have an incentive to increase the volume of those services. Because the payment incentives embedded in product definition shape competition and market performance, it is crucial to design the structure and payments of the delivery system at the appropriate level of aggregation.

For all of these reasons, there has been a growing consensus that competition should be at a level more aggregated than individual services (Porter and Teisberg, 2004), and possibly at the level of fully integrated services (Enthoven and others, 2007). Of course, a move towards more integration must be based on an assumption that it is at least equally as easy for consumers to form judgments about the quality of integrated services as for individual services. And integration may not be superior for everyone; for some individuals, the ability to choose at the disaggregated level is valuable enough to justify the added costs of doing so. The question is whether this group is large enough to justify a policy that supports choice at the individual service level.

The correspondence between competition and product definition in health care means that organisations that are paid to deliver integrated care will do just that, and that they are more likely to combine medical services in an efficient manner. The overall efficiency of any such system of care, of course, is dependent on appropriate rewards for quality or outcomes, to overcome any tendency to under-provide care.

Integration for efficient care

The sources of efficiency gains from better integrating care often seem obvious – for example, the overuse of services that result from misaligned incentives that, at least in principle, could be remedied by integration. Discussions of integration and its benefits in health care make it clear that there are diverse ideas about what integration is, what it might accomplish and what is needed to achieve it. Setting aside for a moment the definitional challenge, one important lesson has been learned in many contexts: efficient performance can only be sustained when both incentives and structure support it; an observation that is true when integration is the means to achieve efficiency.

The path to integration, in turn, is not without cost. Although integrated care delivery can arguably raise quality and lower costs, large integrated care organisations did not get that way without making substantial investments. Unless an organisation that is not already integrated is confident that integration will pay, it will not commit adequate funds and personnel to the effort. Integration often requires costly investments – ranging from the visible, often massive expenditures for electronic health records and other IT, to procedures for better communication, to governance changes – and cannot be undertaken lightly. For example, Kaiser Permanente, the widely-recognised integrated health care delivery organisation, spent \$4 billion as of 2010 for a new electronic health record system to replace one that no longer met its needs. There can be little hope of genuinely integrating care delivery if payment does not encourage either virtual or structural integration of care. Indeed, that is why the physician group and hospital components of Kaiser Permanente evolved in tandem with capitated payment.

Evidence about the effectiveness of care integration is limited. To a great extent, beliefs are driven by the observation that some forms of care are inherently complicated, such as the management of a complex chronic disease, and that multiple providers must care for the patient together. Problems of communication and coordination therefore loom large, in ways that matter less for the treatment of many acute problems such as uncomplicated low back pain, upper respiratory infections and many forms of urgent care. Studying integration is a challenge in part because there is no single definition of integrated

care and its meaning is often context-specific. One meta-analysis examined 15 randomised trials of coordinated care interventions for the treatment of patients with coronary disease, congestive heart failure and diabetes (Peikes and others, 2009). In the included trials, care coordination consisted primarily of nurse-provided education, monitoring, and communication with physicians and patients. The meta-analysis found that the care coordination activities did not consistently reduce hospitalisation rates or expenditures, although there were suggestions that the interventions might have improved some aspects of health.

Although care coordination is often viewed as a product of integration, an integrated organisation need not engage in care coordination of this type, and providers who are not fully integrated can offer coordinated care. What does integration mean, if not care coordination? In the NHS context, integration typically refers to vertical integration, in which hospitals, GPs, and possibly community care providers and others are brought together in a relationship that binds them financially, contractually or organisationally. Often horizontal integration – mergers of like companies or individuals – is thought to be inherently anti-competitive and therefore subject to extra scrutiny. The formation of clinical consortia described in the Coalition Government’s White Paper could represent a form of horizontal integration, for example, although the addition of commissioning responsibilities makes the consortia vertically integrated as well. By forming consortia, groups of physicians might have been able to achieve substantial market power, which they could exercise to the detriment of hospitals and their affiliated physicians. Horizontal integration will ordinarily result in fewer producers of a particular type in a market, giving the producers a more powerful bargaining position. But horizontal integration can also produce genuine economies. For example, capitation seems to lead to more efficient health care delivery, yet payment on a capitated basis requires horizontal integration. A rough estimate of the population needed to be covered for adequate risk pooling in a capitated payment system is around 100,000 people, or approximately the scale of a PCT (and considerably larger than a GP practice).

Vertical integration is often rationalised by a claim that it can increase efficiency, and often the efficiency arguments are persuasive. Vertical integration can be conceived as a means to redefine a product or service, or a production approach, in order to improve efficiency. Even if the efficiency arguments are different, the distinction between

100,000

estimate of the population needed to be covered for adequate risk pooling in a capitated payment system

horizontal and vertical integration is not always simple to apply. Very often, vertical integration requires embedded horizontal integration. A hospital seeking to integrate inpatient, specialty and primary care services will form relationships with a large number of GPs, who will then be linked together as a component of integration with the hospital.

The innovations in integration led by several PCTs may have been spurred by different needs, and may also reflect local conditions. For example, Redbridge PCT developed polyclinics and created programmes crossing primary care and specialised services. As it did so, it took advantage of strong existing relationships between GPs in the area and hospital-based physicians. Many of the GPs had trained at nearby Whipps Cross Hospital, so the specialist physicians and the primary care physicians were able to build upon the trust that comes from longstanding relationships (Moore and O'Meara, 2009).

Previous experience with GP fundholding can be used to justify the views of both proponents and opponents of GP-based commissioning. In the heyday of GP fundholding, many GPs were able to undertake the management components of fundholding and to make this model of practice succeed. But the general view is that only a minority of GPs were comfortable and successful with fundholding. The White Paper proposed that GP consortia would receive what are in effect partial capitation payments. The consortia would use these funds to pay for services that they deliver themselves, and contract for all other services. Under GP fundholding, GPs also bore financial responsibilities for the care of their patients, including care that was provided by other physicians.

These features have not fundamentally changed in the recent revisions to the Coalition Government's health reform proposals. GP fundholding was a mixed success, in which only a minority of GPs embraced their role as managers and felt competent to accept financial risk. The express intent to have GPs join together to form consortia or commissioning groups should mitigate these problems, since large groups of GPs will be better able to pool risk and to obtain the necessary management expertise, either by hiring professional managers or nurturing managers from within their own ranks. Furthermore, the groups are likely to make larger investments in integration if they believe that the government policy will be in place for several years. Then they will find it worthwhile to make the investments in organising

their practices and nurturing or obtaining the management expertise to enable them to operate efficiently as integrated organisations.

Although neither the White Paper nor the Coalition Government's response to the Future Forum report provided details about the ways that the clinical consortia or commissioning groups will interact with hospitals and specialists, apart from ensuring representation of specialists in the consortia, they imply that the groups will continue to pay for such services according to Payment by Results and other specified fees. This all but ensures that such services will not be provided on a fully integrated basis, although the new payment mechanisms appear to have features that will give commissioning groups incentives to limit the use of some services.

Hospitals and commissioning groups will therefore continue to face different incentives regarding the use of hospital services. The groups will want to limit expenditure on hospital and specialty care, while hospitals will typically face the opposite incentives. Large groups may find it advantageous to hire specialists, forming multi-specialty group practices; if they are allowed to do so (the White Paper suggests that this will be possible). Such specialists could coordinate care with GPs, concentrating more effectively on purely specialised aspects of care. Improved communication should also be expected when the physicians in a group all use a single, integrated electronic health record. And a number of compensation models exist for such arrangements, including pure salary payments – similar to the current arrangement for hospital-employed physicians – to partnerships and salaries heavily modified for productivity, which can be defined in terms of both work effort and quality.

Although the door seems to be open to some types of integration, the White Paper explicitly discourages integration between GP consortia and hospitals. For example, Payment by Results or the equivalent will remain in place, so hospitals will continue to be paid on a case-rate basis for inpatient care. One consequence is that hospitals will be reluctant to provide any services that are not included in the case-rate, unless they avoid other costs that the hospital bears (penalties for early readmission, for example). Commissioning groups could provide post-discharge services and handle much follow-up care without the involvement of the hospitals. At other times, the hospital staff will be in a better position to provide such care, particularly following lengthy or complex hospital stays.

The very mechanics of communication with the physicians and nurses providing outpatient care will be greatly improved if there is seamless access to the full medical record, covering both inpatient and outpatient care. Furthermore, hospitals and commissioning groups cannot be expected to make investments in activities that would further integrate care if there are not commensurate rewards; a hospital would therefore be unwilling to contribute to a new inpatient–outpatient programme whose intent was to reduce the need for hospitalisation. And the commissioning groups, which might have the financial incentives to go forward with such a programme, could not succeed if hospital cooperation were necessary.

The Coalition Government’s plans also call for integration between acute and community services. The rationale for such integration would seem to apply to the services that GPs provide or contract for. The commitment to maintain a separate payment system for hospital and specialty care means that integration will not be complete. It will be all the more important, therefore, that under the new reforms innovations which are being developed at several PCTs (Burke, 2010) continue and, if successful, be expanded.

Can competition and integration coexist?

“There is evidence that the quality of care provided by GPs is greater when there is more competition.”

Choice and competitive provision of health care are central to the vision articulated by the Coalition Government. Integration and competition might seem, on the face of it, to be opposite sides of a coin. Market power that comes with large-scale integration has at least the potential to lead to monopolistic behaviour, if not outright monopoly. Yet integration can also lead to lower costs and better quality services; changing the terms of competition in ways that benefit patients. Is it possible to enjoy the efficiencies that come with integration without suffering the drawbacks of monopoly?

The government response to the Future Forum report answers in the affirmative (DH, 2011). And there are reasons to agree with their answer. Although the financial arrangements between individual GP practices and the commissioning groups were still being formulated at the time of writing this report, it appears that the very formation of the groups means that practices will be combined into larger organisations that can purchase services and deliver care in a coordinated fashion. Although this would reduce the number of ‘competitors’ in the English NHS, these changes would not have to lead to the deleterious effects of concentration that would be expected in other contexts. In important respects, choice could increase, not decrease, as long as individuals are able to choose among multiple clinical commissioning groups.

Whether this will ultimately be the case is uncertain. Competition has been limited under the status quo; a person’s choice of GP has been restricted to practices that are accepting new patients and are located within their area of residence. Despite this severe constraint on competition, there is evidence that the quality of care provided by GPs is greater when there is more competition (Pike, 2010). If clinical commissioning groups are indeed limited in geographic extent and individuals have few groups to choose from, competitive pressures on performance will be diminished and might not even improve over the status quo. Furthermore, the lack of choice could mean more dissatisfied patients. Whether a system of commissioning groups offers more choice or less, the most salient harm of monopoly – raising prices above their competitive levels – is largely absent in the NHS, because where prices for services in the NHS are not zero, they are fixed, such as the Payment by Results

“It is not difficult to imagine some clinical commissioning groups becoming very large.”

tariffs. But similarly, if choice does not increase, the putative benefits of increased choice will be lost.

Ironically, the lack of market pricing and other market features is what impedes the ability to reward commissioning groups or providers that gain reputations for quality and service. A successful commissioning group might become enormously popular, if it were able to offer its services to a broader population than its local catchment area. Without the ability to raise prices to capture rewards for their improved care, they could only respond to increased demand by expanding or allowing shortages (in essence, queuing or closing practices to new patients) to develop. Apart from the fortuitous and unusual circumstance in which the administered price is the market price, the most basic role of prices in markets – to equalise the quantity of a product or service supplied and the quantity demanded, eliminating both shortages and surpluses – is sacrificed when prices are administered.

It is not difficult to imagine that some clinical commissioning groups – owing perhaps to reputational advantages, management expertise or other organisational skills – would become very large, spreading across a diverse geography, if they were allowed to do so. At a large enough scale, the advantages of integrating with hospitals would become more apparent. What would be the consequences if they were either allowed to integrate with hospitals formally, or to contract with hospitals for services in arrangements that enabled hospitals and commissioning groups to share financial risk and rewards?

Although vertical integration of this kind can have anti-competitive effects – possibly solidifying market position for the integrated organisation, making it more difficult for smaller groups and independent hospitals to compete effectively – it could also lead to better care. The integrated organisations would need to deliver higher-quality services to succeed. They might attract more skilful managers and deliver genuine efficiencies. And they could not raise prices, since the revenues they received would ultimately be derived from bundled or capitated payments to the groups. Thus patients could expect better services and the care would not cost the NHS more money. One possible outcome of such integration would be the development of several large, regional or national integrated organisations that would compete with one another throughout the

“Clinical commissioning groups might shun some GPs and be highly selective in their choices of partnering hospitals and specialists.”

regions they operate in. Providers could be few in number yet exhibit highly competitive behaviour.

Services could also worsen with integration, particularly if few groups dominate each region and receive little reward for growth, so feel little need to compete. Integration allows the creation and adoption of greater uniformity of care, which by definition reduces variation, inappropriate or not. An integrated care organisation could, for example, develop its own algorithms for care and organise its clinical and business processes accordingly. Such an approach can raise quality, yet uniformity can be excessive. Physician discretion is needed to adapt care in situations that deviate from those that were envisioned when guidelines are written. However, there is little evidence to suggest that care is too uniform today – if anything, variability in care is thought to be evidence of inferior care.

Similarly, in their attempts to improve clinical and operational performance, clinical commissioning groups might shun some GPs and be highly selective in their choices of partnering hospitals and specialists. Although this could be seen as a way to reward performance among all health care providers, selective contracting of this kind would leave some providers out. The White Paper gives the national NHS Commissioning Board the authority to assign GPs to consortia. The Board will feel strong pressure to ensure that no provider is left out, apart from those physicians who are deemed incompetent to practice medicine, and those hospitals that are no longer needed. In this respect, the Board and the commissioning groups may have conflicting aims. The latter will seek GPs who are efficient and highly productive, while the Board may feel the need to protect GPs who provide adequate care but do so inefficiently. If they are unable to select their membership, the GP consortia, reborn as clinical commissioning groups, will seek to use internal incentives – financial or otherwise – to promote productivity among their members.

Incentives for quality and efficiency

Incentive alignment is necessary – indeed a requirement – for sustained performance of any organisation. An integrated provider organisation must be structured appropriately and develop the incentives that will enable it to perform at a high level. The quest for improved health care efficiency has motivated an intense focus on both organisational structure and incentives.

The policy discussion and research about efficient, integrated health care delivery has emphasised organisation. Some organisations appear to be much better equipped to deliver high-quality care than others – a large multi-specialty group practice can better provide team care than can one or more individual physicians or small practices. These capabilities are most important for ongoing, chronic care, as well as care for complex conditions, but these are the conditions that are responsible for much of the disease burden in wealthy nations. In a government-run system such as the NHS, organisation is a matter of design, determined from the top, so getting organisation right is a crucial policy challenge. The NHS, born in the aftermath of war, has a history of varying degrees of central control, particularly for hospital services, yet has conferred extensive decision-making authority on individual and largely autonomous primary care physicians. Throughout its history, reorganisation has been a feature of system reform, often as an accompaniment to changes in payment mechanisms.

In a more market-oriented system, regulation may set limits on allowable forms of organisation of care, but physicians and hospitals have some control over how they work together and how to structure their practices. In a competitive market, organisation is flexible and adaptable, routinely changing as external circumstances change. A change in payment, technological change, and even rises and falls in the prices of factors of production can lead organisations to change as well; think about how the record industry has changed over recent decades, as shops selling vinyl records and later CDs have largely vanished as distribution arms for recorded music. Legislation and government rule-making also induce change, and can do so by a variety of mechanisms. The US Affordable Care Act of 2010 contained elements of both payment changes and direct incentives to alter the ways that physicians interact with one another and with hospitals, in order to create integrated accountable care organisations

“The NHS has continued to include a mix of structural and organisational features of a single corporate/governmental entity.”

(Shortell and Casalino, 2008; Fisher and Shortell, 2010). But the most important levers for change that the law contains are the payment incentives.*

Although the NHS has been subject to the changing aspirations and policies of a succession of governments, it has continued to include a mix of structural and organisational features of a single corporate/governmental entity – the superstructure of the NHS – and a system of incentives embedded largely in payment rules affecting providers. Foundation trusts are by design more independent of direct NHS control than hospital trusts had been, although they are also subject to external oversight through Monitor (an independent regulatory authority) and remain government entities. Health care providers generally are subject to a mix of market-like external incentives – the revenues that accrue from delivering services – and the set of psychological, social and occasionally economic incentives that are characteristics of large organisations. In the NHS, as in most economic settings, the incentives can be overt or implicit, unplanned or the product of conscious design. To understand both the evolution of the NHS and the likely outcomes of the current reforms, it is useful to consider how structure and incentives interact.

Performance incentives in medicine

Because they make most health care allocation decisions, the behaviour of physicians, other health care providers and patients is fundamental to health system efficiency. That is why it is so important to put in place incentives to promote efficiency. In the NHS, a system that insulates patients from financial costs, price cannot be used as an incentive to influence individual patient behaviour. But provider incentives are used and there is ample evidence that financial incentives influence provider behaviour.

The history of Kaiser Permanente is a case in point, with particular relevance to the NHS. Unlike most health insurance arrangements in the US, from the beginning, Kaiser Permanente has paid providers on a capitated basis – providers receive a fixed fee for

* Whether the payment incentives are sufficient is another matter, as the debate over implementation has revealed.

“Members of health maintenance organisations used hospital care nearly a third less than individuals enrolled in conventional fee-for-service insurance.”

the provision of all care to a defined population, rather than payments for each unit of service. Thus they had strong incentives to limit the use of expensive resources and to avoid unnecessary services.

The most celebrated controlled trial in health services research, the Rand Health Insurance Experiment, showed that members of health maintenance organisations, which were paid on a capitated basis, used hospital care nearly a third less than individuals enrolled in conventional fee-for-service insurance (Newhouse, 1993). Because the Rand study was a randomised trial, it overcame the criticism levelled at nearly all other studies of different forms of insurance: individuals select insurers, rather than being randomly assigned, so differences in utilisation and outcomes might reflect unobserved differences in the characteristics of individuals who chose different health plans, not the performance of the plans themselves.

The Kaiser Permanente experience illustrates the linkage between structure and incentives; the entire organisation (or more accurately, set of organisations) is set up to support care that is efficient under capitated (per-patient, rather than episodic) payment. Without distinctive capabilities that allow them to provide care in a coordinated fashion, efficient capitated care is beyond the reach of independent physicians and hospitals. A provider must operate at sufficient scale to be able to maintain the management tools and capabilities needed to track resource use and patient outcomes. It must also be large enough to bear the financial risk of caring for the occasional patient with an extremely costly illness, such as organ failure requiring transplantation, or a cancer requiring prolonged treatment with expensive biological compounds. And pooling of financial risks and rewards is usually necessary to ensure that providers of different components of care will work together effectively to achieve common objectives. Thus organisation and structure, particularly in a lightly regulated, competitive system, must match incentives that are embedded in the rules governing payments for care.

This observation, of course, also applies to systems that are highly regulated or established by governments. However, in such a system an inefficient, uncompetitive provider will not inevitably be forced out of the market. Politics often trump economics in determining survival in system such as this.

“With inadequate risk adjustment, a physician or health care organisation can achieve better health outcomes simply by selecting healthier patients.”

When we think about performance incentives, it is useful to distinguish between those that originate outside an organisation – for example, the payment rules – and the set of incentives that the organisation uses to promote desired behaviour within – in other words, the influences on members or workers. Internal incentives can include bonuses, ‘employee of the month’ commendations and promotion policies. Internal incentives are closely bound up with structure; in a large hierarchical organisation with a great deal of central control, for example, it can be difficult to gain accurate information and to encourage innovation at the widely dispersed lower levels of the organisation.

The motivating goals are not always stated explicitly, or they may be expressed in narrow terms that only bear a loose relationship to the ultimate goals. At the system level, quality improvement is often the stated goal. But usually it is not the ultimate goal. For example, despite the occasional use of language that suggests otherwise, quality at any price is not the goal. That is because quality standards usually incorporate, if only implicitly, notions of cost-effectiveness or productivity. Productivity, in turn, may be expressed in terms of outcomes obtained from a fixed expenditure, or, much less commonly, minimising the expenditures needed to achieve defined health goals. Seldom is the real goal the pursuit of improved health outcomes without any notion of value or recognition of the costs of achieving the outcomes. And there are other reasons not to focus exclusively on outcomes achieved.

A one-dimensional goal such as outcome improvement can lead to behaviour we would seek to deter. For example, with inadequate risk adjustment, a physician or health care organisation can achieve better health outcomes simply by selecting healthier patients. It pays to be explicit about the ultimate goal, even when there is not a complete consensus that it is the right goal. A strong focus on a related goal can too readily lead to the wrong results.

The design of incentives for business and government organisations has been analysed and debated for many years (Milgrom and Roberts, 1992; Laffont and Martimort, 2001). At least in business, articulating goals is often straightforward. Public corporations, for example, exist to maximise shareholder value. A non-profit, non-governmental organisation will often have a broad goal that is clearly stated, even if there is no counterpart to a stock price to assess whether or how well a

goal is achieved. Kiva, an international microfinance non-profit organisation, states that its mission is ‘to connect people through lending to alleviate poverty’. For Oxfam International, the goal is ‘to find lasting solutions to poverty and injustice’. Articulating a goal requires discussions about the most basic values and purposes of any organisation. It is fundamentally about organisational identity.

At one level, a corporation seeks to maximise profits. But organisational identity may relegate profits to a less prominent role. For an industry such as IT, for example, innovation may be viewed as more important than many measures of profits. The choice between making profit maximisation the only goal and making innovation a goal in itself can shape the direction of the firm. These distinct goals can be reconciled by a belief that measures of innovation are more accurate guides than short-term profit measures to long-term profitability.

Settling on a goal is only the first step in implementing incentives. It is also necessary to identify the observable behaviours that the incentives are designed to encourage. A health organisation might agree that it would like to improve quality of care, working from a fixed budget. But to have the greatest measurable impact, it might well decide to target its quality improvement activities at the conditions and patients most likely to benefit from improved care. That is why health care providers so often focus on chronic diseases, in which relatively simple and inexpensive interventions can lead to better outcomes. That is also why their goals are more specific than general quality improvement – they focus on outcomes such as better blood sugar and blood pressure control among diabetic patients.

This somewhat more specific focus is still a step removed from implementation. Implementation requires measures that form the basis for rewards and penalties. The magnitude of the challenge is apparent from the large literature describing existing quality metrics, the characteristics that they should possess and their correlations with outcomes such as mortality rates (Mant and Hicks, 1995; Iezzoni, 1997; Thomas and Hofer, 1999; Davies and others, 2001; Balk and others, 2002; McGlynn, 2003; Dimick and others, 2004; Bradley and others, 2006; Werner and Bradlow, 2006; Jha and others, 2007). Ease of measurement is a practical requirement for any quality metric used as a basis for incentives, but any difference between the metric and the ultimate goal, when a powerful incentive is in place, can be

counterproductive. Large incentives induce behavioural change consistently and, at times, dramatically, so it is important to ensure that the incentives induce the desired behaviour. However, the criteria or targets are – nearly always – proxy measures for the desired behaviour, not the behaviour itself. In a naturalistic setting – perhaps observed behaviour before the performance incentives are put in place – the metric and the desired outcome may be tightly correlated, but the correlation can weaken substantially when the metric is used to assign rewards.

‘Teaching to the test’ is an example. A test in a subject such as European History, particularly when given unexpectedly, may be a very accurate gauge of a student’s knowledge of the subject. But performance on a more typical examination, when the questions are identified beforehand and students can formulate specific strategies to do well on the test, reflect test performance strategies as well as mastery of the content. Thus the outcomes that are measured and the targets chosen should be intrinsic to the desired outcome, rather than measures of related phenomena.

For reasons that sceptical clinicians will readily offer, this ideal is not easily achieved. It is easy to see why: few quality measures are direct outcome measures. Process-based performance criteria are used far more often; they are attractive precisely because process adherence is much simpler to observe and to apply (Cromwell and others, 2011). Outcome measures, in contrast, are beset with obstacles. Successful use of an outcome measure, if it is to be used to promote better patient care, requires distinguishing the contribution of the clinician to the outcome. Doing so requires adjustment for the underlying condition of the patient being treated; separating the contributions of the clinician from those of other care providers and from patient behaviour itself; choosing an appropriate time frame for the measurement; and finding conditions in which the random variation is not so large that it is difficult to discern the provider’s contribution to the outcome.

Consider the challenges in developing performance criteria based on readily observed complications of diabetes, such as the development of renal failure or heart disease. Because both complications take many years to develop, they have little relevance for a very long time, and by then it is seldom feasible to attribute outcomes to any one clinician. Their occurrence is influenced by patient behaviour,

“According to one review, performance incentives must account for five per cent or more of a physician’s income to induce desired behaviour change.”

unobserved differences in health and other circumstances that the clinician does not control. Furthermore, unless the diagnosis of such conditions is governed by a fixed protocol, variation in complication rates may simply reflect differences in diagnostic practices. This is a particular issue for providers who care for patients living under conditions of social deprivation, who tend to have worse outcomes even when adjusted for observed health characteristics.

Performance incentives have been tested extensively in medical care, sometimes under controlled circumstances, with mixed and, perhaps, disappointing results. In the US, these incentives have been incorporated into pay-for-performance (P4P) programmes. In the face of strong physician resistance to the programmes, the financial incentives have often been small (for example, less than one per cent of the fees is subject to performance rewards or penalties). Perhaps unsurprisingly, the effects of these small incentives can be difficult to discern and sometimes work in the wrong directions. According to one review, performance incentives must account for five per cent or more of a physician’s income to induce desired behaviour change (Young and others, 2007).

Many performance incentive programmes concentrate on the first steps of performance improvement, requiring only that physicians or hospitals report whether a service was rendered. They recognise that for many providers even the basic requirement to report whether services were performed or to record the value of a lab test can be a challenge. It is not unusual to reward a physician simply for reporting a recent glycated haemoglobin value for a patient, even if the level did not fall into a desirable range. In some programmes, the ability to show that the test had been performed at recommended intervals – regardless of the ability to report the value – has been sufficient to earn a performance reward.

Often the ability to report whether a test was performed requires a substantial investment, at least to do so efficiently; with an electronic medical record system the effort involved may be trivial, while reviewing paper charts can require too much staff time to warrant the effort, unless the payoff is large. But installation of an electronic medical record system requires dedicating funding to software and services, as well as substantial staff time. The investment is typically too large to be justified solely by the desire to earn performance incentives. The case for investing in the electronic medical record

system weakens when there is doubt about whether the performance incentives will remain in their current form for long, and whether the investment will soon be rendered obsolete by new technology or new standards.

Despite their growing popularity, the support for performance incentives is often indirect. Direct trials of the effects of performance incentives on the quality of care have provided only weak support. In a review of studies of the effects of performance incentives on the quality of care that was published in 2006, Petersen and colleagues found 17 studies that met their inclusion criteria, of which 14 examined programmes that rewarded processes rather than outcomes of care (Petersen and others, 2006). In some cases, the performance incentives led to better documentation (for example, of delivery of preventive care) without proof that they increased the actual utilisation of the services. Only five of the 14 studies had unambiguously positive findings, but the variation in study designs, incentive designs, domains (nursing home care as well as different medical settings) and in the magnitude of the performance rewards made it difficult to draw any firm or broad conclusions. In fact, if the studies had been subjected to the tests of heterogeneity that are commonly performed in meta-analyses, it is unlikely that they could have been pooled in any meaningful way.

The Premier Hospital Quality Incentive Demonstration, a large Medicare project which started in 2003, used both financial rewards and public reporting of results as incentives to promote improvements in process-based measures of care quality (Lindenauer and others, 2007). Compared to a set of matched hospitals that engaged in public reporting, the hospitals enrolled in the demonstration were more likely to achieve the performance targets, even though the financial incentives were modest – a two per cent premium over the usual payment for hospitals achieving the top performance decile, and a one per cent premium for those reaching the second decile. The performance criteria were based on well-accepted standards; for example, for the treatment of myocardial infarction, the measures included the percentage of patients receiving such treatments as aspirin on arrival, beta blocker on arrival, ACE (angiotensin converting enzyme) inhibitor or ARB (angiotensin receptor blocker) for left ventricular dysfunction, aspirin on discharge, and a beta blocker on discharge.

In a recent study of the same demonstration project, Werner confirmed that the quality improvements were maintained for the first three to four years (Werner and others, 2011). About five years after the start of the project, however, the matched hospitals did nearly as well. As expected, the hospitals eligible for the largest incentives improved the most; hospitals in non-competitive markets, which faced little external pressure to improve performance, had larger responses to the performance incentives than hospitals that faced more intense competition for patients. Without competition, there was no need to compete over quality, while the cash incentives for quality improvement would still matter. Perhaps the most surprising aspect of this study was that the incentives had much effect at all, since the incentives were small enough and covered so few clinical conditions that the average size of the incentive was in the order of only \$12,000 per hospital.

Other authors have similarly questioned the magnitude and duration of the effects of financial incentives (Rosenthal and Frank, 2006; Rosenthal and others, 2006). But they have also noted the weaknesses of the primary studies: randomised controls are typically absent, the incentives are often too small to exert strong influence, and with non-randomised designs the controls are also subject to changing incentives during the period of observation.

Many studies do not even have contemporaneous controls. Furthermore, studies that focus on Medicare typically have no data on non-Medicare patients. Any physician or hospital contemplating an investment in an electronic health record system, for example, will assess the benefits that result from the care of all their patients, not only Medicare beneficiaries. Often payments from commercial insurers will contribute more than Medicare payments to the revenues that result from the investment. This omission, therefore, can impede efforts to find the effects of one payer's practices on the behaviour of health care providers.

Some critics object to the principle of using financial incentives to improve quality or other aspects of clinical performance. They believe that professional norms and a commitment to patient wellbeing should motivate all such efforts. But more often, concern extends from unease about the measure itself. Often the measure's relation to the desired outcome is weak or incomplete. And the design of the incentives raises many technical issues. For example, programmes to

promote quality improvement often reward physicians for achieving quality targets, but some experts on performance incentives in medicine argue that rewards should be linked to improvement rather than to the attained level of quality. They base their arguments on a belief – often grounded in experience – that an absolute quality standard would be too easy for some providers to obtain, while for others it would be unachievable (Rosenthal and others, 2004).

Basing rewards on improvement rather than achieved quality may lead to better quality, but can only be justified on the basis of distortions in the market for medical services. After all, no consumer of an ordinary product or service would pay a premium for a ‘most-improved’ item if it remained inferior to the alternatives. And poor quality providers would lose business. But in medical care quality is typically difficult to measure, and patients have little access to good measures of provider quality. Furthermore, small rewards for quality offer little incentive for providers.

In the US Medicare programme, physician resistance to pay-for-performance meant that the amount of money at stake was limited to less than two per cent, as in the Premier demonstration project. Even in one of the most ambitious pay-for-performance demonstration projects, the Physician Group Practice Demonstration Project, physician groups were allowed to retain most of the savings that Medicare would receive from their more efficient practice, but the savings were capped at five per cent and were expected to be substantially smaller (Leavitt, 2006).

If the rewards are small, the payoff to even modest investments in quality improvement may be insufficient to change behaviour, so the temptation is to set a low bar or to base the payment on the amount of improvement. These are essentially short-term strategies, because they are inefficient and unsustainable; they reward change rather than attained quality, with larger potential rewards to under-performers and implicit penalties for providers that are already doing well.

Limits of financial incentives

In medicine, financial incentives are often small because they represent a compromise between competing interests. During policy debates about Medicare pay-for-performance initiatives, physicians let it be known that they would only welcome such programmes if they were

“Professionalism and other non-financial influences on behaviour can be powerful motivations.”

used to increase total compensation. Members of the American Medical Association opposed a large pay-for-performance programme because they assumed that physicians would be subject to reductions in compensation if they failed to meet prescribed targets. They did not believe that the performance payments would be added to an existing baseline payment. But rewarding performance, without any payment reductions for those who fail to meet performance targets, requires incremental funds, an unattractive option for policy-makers trying to reduce budget deficits.

Small performance rewards might not only be inadequate to cover the fixed costs of complying with quality improvement targets; they could also be counterproductive. Within medicine there has long been a strong undercurrent of antipathy to the use of financial rewards as a means of promoting better medical care. According to this view, it is demeaning and fundamentally unprofessional, if not unethical, to receive a financial reward for delivering better quality health care, which physicians should do anyway. Professionalism and other non-financial influences on behaviour, in fact, can be powerful motivations, and objections to the use of financial incentives, based on the possibility that they will be counterproductive, have some experimental support.

As Gneezy and Rustichini (2000b) observed – following an extensive psychological review (Deci and others, 1999) – a financial incentive can induce the opposite behaviour of what was intended. This paradox occurs because a monetary reward is a form of extrinsic motivation that can undermine an intrinsic motivation such as a sense of duty. They describe a controlled experiment conducted in an Israeli day care facility, in which a fine was imposed on parents who were tardy when the time came to pick up their children at the end of the day (Gneezy and Rustichini, 2000a). After the fine was imposed, the rate of late pick-ups increased among the group subject to the fine. According to one explanation, the guilt or opprobrium they felt when they were late was now expiated – assuaged by the knowledge that they would be paying a fine.

According to these authors, there is a ‘W’-shaped relationship between financial incentives and behaviour change. A fine or reward that is too small has paradoxical effects, as seen in the Israeli day care experiment, because it nullifies intrinsic motivation without being powerful enough to induce desired behaviour. This

does not mean that a fine or a reward will never be effective; it just needs to be large enough to overcome the loss of the intrinsic motivation.

Organisational change requires individual behavioural change, so observations such as these suggest that strategies to promote better care in medical settings will be expensive if they are based solely on extrinsic motivation. Furthermore, it is entirely possible that some physicians, if not entire specialties, are more susceptible to extrinsic motivation, while others are driven largely by intrinsic motivation. For many surgeons, for example, the desire to have a successful operative outcome will often have much greater force than a modest financial reward or even formal recognition of surgical excellence. For other physicians, and perhaps for other outcomes that they find to be less important or compelling, financial rewards and other extrinsic motivations may carry greater force. Thus the most effective strategies will likely include a mix of incentives.

The evolving structure of care organisation and performance incentives in the English NHS

Performance incentives in the NHS have been adopted and have evolved in parallel with changes in the organisations that commission, oversee and deliver care. In a market economy, organisations grow and adapt in response to external incentives, but law, regulation and government policy shape organisations in a government-run sector. In the NHS, planned organisational change has been a necessary complement to changing incentives. This is evident with such major payment innovations of recent years as Payment by Results and the new GP contract. The organisation of hospitals, local commissioning authorities and GP practices had to change as well.

In the hospital and secondary care sector, for example, there was a long-term shift from fixed budgets to revenues directly determined by the value of services rendered. This was made possible in part by the creation of NHS trusts, which were introduced in 1991 with the Working for Patients set of reforms. This organisational innovation was intended to ensure that inpatient and specialty services would be delivered with greater efficiency, promoted by organisational autonomy. The trusts had distinct corporate identities and generated revenue by billing for services provided.

“Commissioning is essentially an insurance function.”

The establishment of locally owned foundation trusts in 2003 was a further step in the evolution of these hospital-based organisations. Foundation trusts were granted more autonomy than hospital trusts, with a structure akin to that of a private corporation, with less direct control from government, free from direct oversight of SHAs, and with flexibility in managing finances and accessing capital. In particular, although they needed to be financially viable, they were not required to break even each year and could retain surpluses.

To earn the foundation trust designation, a hospital trust needed to demonstrate management capabilities such as the ability to track costs and finances. Monitor, an independent regulatory authority, oversees foundation trusts. The capabilities that were necessary to become a foundation trust were those that would be needed to respond quickly and effectively to changes in financial incentives, such as those represented by Payment by Results.

As of June 2011, 137 of 247 hospital trusts were foundation trusts. The White Paper expressed the intent to convert all hospital trusts into foundation trusts, and to lessen central control even more by reshaping the governance of foundation trusts so that they would become free-standing ‘social enterprises’. The Coalition Government announced in June 2011 that it would not pursue the foundation trust conversions as quickly and aggressively as it had planned, but it did not abandon the goal. The shift to greater autonomy of hospital trusts is likely to enhance the effects of financial incentives: the foundation trusts will have greater freedom to pursue alternative policies to improve performance; there will be greater ability to marshal financial and human resources to improve productivity, especially greater freedom to obtain capital; and better financial reporting and management skills can facilitate improvements in operational performance.

PCTs may not have been created to respond to a specific set of performance incentives, but they have similarly evolved to develop considerable autonomy as well as local responsiveness in order to promote efficiency in the delivery of health services. This seems to have been the intent of World Class Commissioning. Commissioning is essentially an insurance function, albeit one with greater similarities to American managed care, than to simple indemnity or fee-for-service insurance. The disenchantment with PCTs that led to

the proposal to abolish them bears remarkable similarities to the opprobrium that has often been directed against managed care in the US. Some of the criticisms may be well founded – some PCTs may not be well managed, and many are overmatched by both GP practices and hospital trusts. Indeed, ongoing restructuring of commissioning reflects the pursuit of more professional contract management, better management of population risk and improved demand management. The creation of PCTs as the local contract holders gave them, at least in theory, the authority to combine the provisions in the new GP contract with Payment by Results to procure a better mix of services for populations. However, there has been considerable uncertainty about how much autonomy PCTs have to organise relationships with GPs and with hospitals differently, and they seem to operate with less flexibility and independence than, for example, foundation trusts.

Whether the apparent failure to succeed is a result of management weakness or more systemic problems, the US experience confirms the difficulty of the role that PCTs play: the physicians and hospitals are not employees; there are few levers to manage utilisation; competition among GPs is geographically limited (and, notwithstanding Choose and Book, secondary services do not yet appear to be subject to intense competitive pressures); and patients are more likely to trust the recommendations of their GP or specialist than those of the PCT. The intermediary role represented by commissioning is inherently challenging. Clinical commissioning groups will be subject to the same pressures and difficulties, and possibly similar incentives.

Recent national developments that have had an impact on incentives

During the Labour Government of the first decade of the 21st century, the main national incentives of the ‘system reform’ period of the NHS in England have been the introduction of Payment by Results in 2002 and the new GP contract in 2004. Alongside these main incentives, associated with the financial flows of the system, a new consultant contract was adopted in 2003.

Payment by Results is essentially a case-rate payment system for inpatient and secondary services. It was introduced in order to improve several aspects of the performance of the NHS. Specifically, it was intended to improve the efficiency of inpatient care, to increase

transparency for purchasers (in other words, the PCTs), to provide a source of performance-based revenues that could be used to finance appropriate expansion in hospital capacity, to reduce waiting times for procedures, and to give patients more choices in hospital care. At the outset, only 25 surgical procedures were included in Payment by Results, but a series of revisions greatly expanded its scope. Today, it covers most acute care delivered by hospital trusts, including outpatient services, diagnostics and unscheduled care.

Much like Medicare's Prospective Payment System in the US, Payment by Results relies upon a 'grouper' that assigns each hospital admission to a Healthcare Resource Group (HRG), which in turn has an associated tariff or case-rate payment. The HRG payment rate is derived from a set of reference costs that are in turn based on the resources expended on behalf of a nationally representative sample of patients. In other words, the HRG rate is a case-rate payment that is designed to be representative of the average costs expended on patients throughout the entire NHS.

With revenues now dependent on case-based payments, and a clear reason to limit underlying costs, Payment by Results provided strong rewards for better management, especially in foundation trusts. It seems to have succeeded at increasing transparency and strengthening financial management among both hospitals and commissioners. What of its direct incentive effects? The literature on the introduction of DRGs (diagnosis-related groups) as part of the Medicare Prospective Payment System in the US hinted at the future experience of the NHS. As would be expected, activities that had more favourable margins – that were more profitable – experienced growth. In particular, the rate of elective procedures grew. From the point of view of the government, an increase in elective procedures was not necessarily objectionable, since it might have been necessary for the successful effort to reduce queues for elective procedures. Of course, that outcome was not a foregone conclusion, since greater demand for elective procedures, unless accompanied by still greater increases in throughput, would actually lengthen queues.

A report by the Audit Commission (2005) identified setting the tariff – the case-rates – under Payment by Results as a challenge. The report also raised concerns about potential 'gaming', for example, from changes in diagnostic coding to procure more favourable payments (upcoding) and keeping patients for at least 48 hours to receive

“Integration of GPs and hospitals with shared risk was widely thought to be discouraged.”

higher payments and transferring shortly thereafter. Although these were described primarily as potential rather than known problems, some PCTs had reported cases of suspected gaming.

Payment by Results also facilitated private provision of health care. By offering clarity, especially about revenues, it allowed non-NHS providers to more seriously consider offering medical services. The defined payment schedule made it possible for independent sector (private) providers to define business plans that could in some cases support entry into the NHS. Specific inducements – including the protected contract, which could provide guaranteed income for new provider entrants – were used to boost private provision of care, and there was an increase in (predominantly surgical) private provision between 2002 and 2005. However, this effect may have been transient, since there was a subsequent fall-off in the number of private providers. Many of the independent sector treatment centres of this period have since withdrawn from the market.

The foundation trusts, with their greater autonomy and, presumably, better developed management capabilities, were well positioned to adapt to the features of Payment by Results. This has led to an imbalance in local power – with foundation trusts able to attract stronger managers with higher pay. The requirement for commissioners to pay hospitals at the prescribed Payment by Results tariff for all work done has meant that the hospitals have increased their revenues substantially. PCTs were charged with managing demand, an essential element in a market for services in which care providers had strong incentives to generate greater utilisation of their services. But the PCTs turned out to be overpowered by the acute trusts in this regard, lacking the levers to limit demand for hospital services. For example, the GPs who referred patients to hospitals were not at financial risk for such care, and there was no direct mechanism by which a PCT could impose restrictions on the use of hospital services. An obvious solution – integration of GPs and hospitals with shared risk – was widely thought to be discouraged, if not overtly prohibited. Indeed, with few exceptions, GPs were not allowed to send patients to preferred providers since this would conflict with competition policy.

The designation of preferred providers is a form of selective contracting that can appear to be anti-competitive but can, in some circumstances, increase patient choice or improve other aspects of patient wellbeing.

A GP and a hospital or set of specialists might engage in such contracting as a means of lowering costs (for example, because of better communication and other efficiencies that result from extended, close working relationships) or quality assurance (only providers who met quality standards are eligible for a preferred contract). A prohibition of preferred provider designations can, under these circumstances, lead to worsened services for patients. Nevertheless, as in any other form of increased integration, preferential contracting has at least the potential to be used as a way to engage in virtual vertical and horizontal integration, leading to the possibility of market concentration and monopoly power as discussed earlier.

The implementation of Payment by Results by hospital trusts appears to have functioned as designed, if not precisely as intended. Overall, according to a 2008 Audit Commission report, the effects were modest; for example, although elective admissions increased somewhat, they did not increase as a proportion of all hospitalisations (Audit Commission, 2008). And in some respects it echoed the American experience from the early 1980s when Medicare shifted from pure fee-for-service to a case-based payment system for hospital care (Kahn and others, 1990; Sloan and others, 1988; Draper and others, 1990; Coulam and Gaumer, 1991). At least initially, the adoption of Medicare's prospective payment system led to a decline in lengths of hospital stay, a lower hospital admission rate and a decrease in Medicare hospital expenditure. The expenditure decrease, however, was partially offset by an increase in spending on outpatient and other medical services (Lave, 1989).

The Quality and Outcomes Framework and local incentive schemes

The new GP contract was introduced in April 2004, replacing the 'red book' schedule of payments for GP services with a number of new features. The changes introduced were, for the most part, highly favourable to GPs. For example, primary care organisations (PCOs) assumed the responsibility to offer care on nights and weekends, relieving many GPs of this responsibility. Among the changes was the introduction of QOF. The incentives introduced by QOF were much stronger than those embedded in most of the highly publicised pay-for-performance programmes in the US, including those that were part of official Medicare demonstration projects.

“The ‘prescribing incentive scheme’ has been a powerful driver to increase the use of generic drugs.”

QOF included a variety of performance incentives and criteria; many were very simple, such as recording the percentage of patients with coronary heart disease who were questioned about smoking within the last 15 months, or the ability of a practice to maintain a register of patients with hypertension. As described below, QOF targets were met quickly.

Local incentives are embedded in Commissioning for Quality and Innovation (CQUIN), a framework introduced in 2009 to enable local commissioners to develop quality improvement goals and to offer financial rewards for their achievement. CQUIN ‘is intended to ensure contracts with providers include clear and agreed plans for achieving higher levels of quality by allowing PCTs to link a specific modest proportion of providers’ contract income to the achievement of locally agreed goals’ (NHS Institute for Innovation and Improvement, 2011). It was based in part on the Premier demonstration in the US Medicare programme (Maynard and Bloor, 2010). By definition, CQUIN schemes are varied. Furthermore, the total amount of payment at risk is small – about one per cent – but some of the programmes appear to have had a substantial impact on practice.

One common example is a ‘prescribing incentive scheme’ whereby each practice is given a nominal prescribing budget, typically based on last year’s activity and adjusted for inflation and changes to the formulary. The practice is then given frequent information on utilisation, along with some support – usually in the form of a pharmaceutical adviser – and is allowed to keep a proportion of any savings generated below the budget. This has been a powerful driver to increase the use of generic drugs, and to make GPs more mindful of prescribing practice.

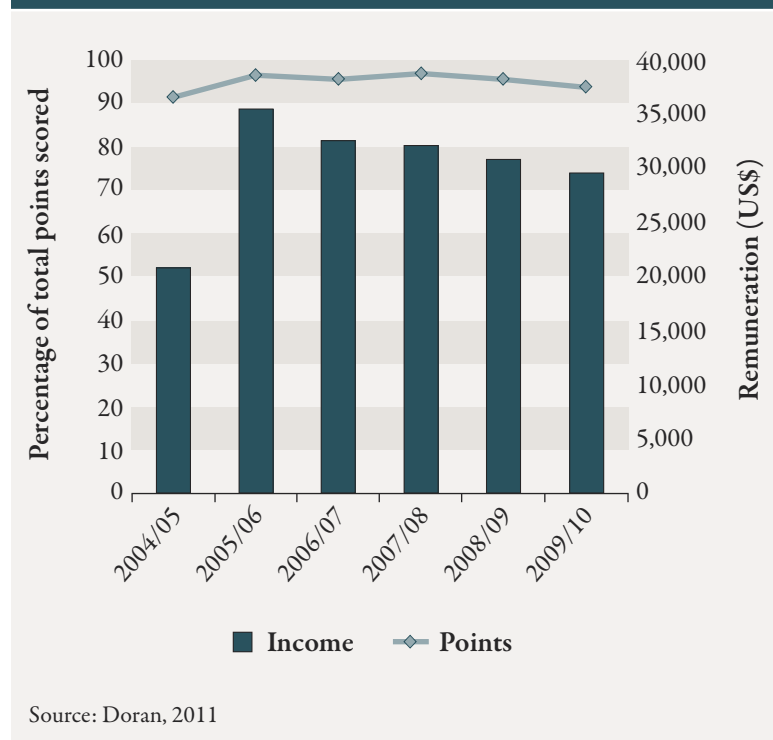
Other local schemes have included referrals management, care management and the introduction of services to avoid hospital admission or to speed up discharge. These ‘demand management’ initiatives met with only patchy success, not least because they run counter to the core financial incentives of the hospital trusts.

These programmes and other aspects of the new contract significantly raised the income of GP practices. This has brought average GP income in the NHS broadly to parity with hospital consultants, with the ability for entrepreneurial GPs to achieve incomes in line with proceduralists within private practices.

Effects of incentives in the English NHS

The NHS experience confirms that large incentives have effects. But beyond that simple statement, the role of financial incentives, in particular, has been controversial. Under QOF, a GP could augment his or her income substantially. After QOF was put into effect there was a remarkable increase in the percentage of practices that came into compliance with the desired behaviours. As Figure 2 (below) shows, in 2004/05, on average, practitioners reached about 90 per cent of the available points and about \$20,000 in incentive payments; in the following year, the average payment rose to about \$35,000, and remained at similarly high levels with modest declines in subsequent years.

Figure 2: GP performance under the Quality and Outcomes Framework



“The effects of incentives are inherently difficult to measure.”

Similarly, Payment by Results was viewed almost immediately as being successful in altering the behaviour of NHS trusts, as hospitals were rewarded for activity and, particularly if they were foundation trusts, penalised for operating at high cost.

Observations like these are suggestive, yet do not by themselves prove that the financial incentives embedded in such programmes lead to sustained benefits. Evaluations of incentive schemes have pointed out that changes in behaviour occur without the incentives as well, that the effects may only be sustained as long as the incentives are in place, and that experiences with incentive programmes may not generalise to other settings (Campbell and others, 2007; Campbell and others, 2009), as has been found in the US. But the effects of incentives are inherently difficult to measure, if only because the motivation for applying the incentives would affect care even if the incentives had not been adopted. The goal of incentives is to accelerate and intensify the dissemination of preferred forms of care, or better outcomes, and even if they are successful they will be competing with other approaches to improving care. Without a randomised comparison, it would be difficult to tease out the effects of specific incentive programmes. Thus the central question is not whether financial incentives induce behaviour change; they do, if they are large enough. The more relevant question is whether alternative approaches can improve health more, or do so at lower cost. Here strong evidence is likely to remain elusive, without new randomised studies or natural experiments.

Equally uncertain is the role of non-financial incentives. Professionalism, social norms, recognition and a host of other approaches to motivating desirable behaviour can undoubtedly be effective and need not require money. Very little is known about how financial and non-financial incentives can be used together in health care settings, and whether it is possible to avoid the problems that arise when extrinsic motivations, which are not necessarily financial, vitiate the benefits of intrinsic incentives. For example, would the regular expectation of praise for a job well done, or payment for the same, cause physicians to shift their efforts from the less observable aspects of high-quality care towards those that others can easily see? And, if so, how can one judge the relative importance of these distinct aspects of a physician's work?

Indeed, the most important difference between alternative approaches to awards may not be so much about whether they are

“The most important difference between alternative approaches to awards may be whether the criteria are clearly defined.”

cash payments as in QOF, but whether the criteria for the awards are clearly defined. To administer a pay-for-performance programme, it is necessary to apply criteria that are readily defined and quantified, and to impose numeric formulas for the relationship between the reward and the value on the performance criterion. But promotions in organisations, though they are often based on defined criteria, are never determined as mechanically and they are often vague. Why?

Vagueness can be an asset, if highly specific or rigid criteria are either incomplete or imperfectly related to the characteristics that really matter. To enhance the productivity of a hospital-based physician, for example, the hospital might develop rewards based on the number of patient visits per month, adjusted for complexity. But this measure of volume might be insensitive to the quality of care the physician delivered. An incentive, financial or otherwise, that offered rewards based on this measure of productivity, would promote volume rather than quality. If it is difficult to quantify a dimension of care such as quality or patient satisfaction, and ‘fuzzy’ measures such as reputation are closely related to the desired aspects of performance, a highly structured incentive approach might be less successful than a relatively vague one. Incentives that are highly structured and specific are more vulnerable to ‘gaming’ and less able to reward meritorious behaviour that doesn’t readily lend itself to quantification.

These potential advantages of discretionary approaches to performance rewards should not blind us to their drawbacks. Explicit performance incentives are popular because they work in measurable ways. Less well-defined approaches allow not only for appropriate discretion, but also for arbitrariness and for the influence of pettiness, favouritism and a range of emotions that should have no place in determining rewards. That is why the quest to discover the best blend of incentives, both financial and non-financial – based on a combination of well-defined and vague performance criteria – is unlikely to end any time soon.

Health reform in England and the US

This discussion has emphasised the importance of both organisation and incentives, and the potential for competition to improve health care productivity. But it has also described weaknesses in the evidence base supporting particular strategies in each of these areas. The available research permits only limited claims about the superiority of any one approach to health care delivery over others. Recognising these limitations, what are the implications of these observations for health reform efforts in England and the US?

If there is one overarching lesson from the experience of the English and the US health reform efforts of the last two years, it is that policy proposals can differ dramatically from the legislation that grows out of them. Reform proposals in both countries had to be amended, often in fundamental ways, as they were debated and criticised. The vision of the White Paper, for example, was ambitious, but it was criticised from within and outside the Coalition Government, and there appeared to be considerable resistance within the Department of Health. Even GPs, who were given much greater authority and influence, reacted tepidly to the proposed changes. Central to that vision was the idea that enhanced competition, including an expanded role for private health care delivery, along with changes in organisation, would lead to more efficient care. Taken to its logical conclusion, it might have led to greater integration, with fewer entities providing care but doing so at larger scale and, ideally, greater efficiency.

Another goal was to offer patients greater choice. But public questioning and debates raised basic questions about choice. How is it to be defined, and how can we know when there is more choice? Effective choice – giving patients real alternatives that are relevant to their needs and that are attractive to them – does not depend solely on the number of independent providers. With fewer organisations, but well-managed ones, quality and efficiency improvements would respond better to patient needs and desires. Furthermore, large integrated groups typically give patients a choice among individual physicians within the group, even when they discourage the use of physicians outside the group. This contrasts with the limited choice of GPs that now characterises the status quo in the NHS.

These questions could be readily anticipated from the White Paper itself. In the document, competition was sharply circumscribed, with no freedom to compete for primary care across geographic boundaries. And following the public commentary on the proposed legislation during the listening exercise, the Coalition Government announced changes in the planned reforms, largely in line with the recommendations of the NHS Future Forum. The revised approach downplayed the role of competition, changing Monitor's charge from promoting competition to promoting quality. In the words of the Future Forum report, 'the bill should be changed to be very clear that Monitor's primary duty is not to promote competition, but to ensure the best care for patients. As part of this, they must support the delivery of integrated care' (DH, 2011: p25). The Future Forum also called for specialty representation on the GP consortia, which was part of the rationale for renaming them clinical commissioning groups (DH, 2011). The Future Forum also expressed scepticism about the readiness of many areas to form consortia with the requisite competence in time to meet the proposed deadlines. Overall, the changes in response to the Future Forum report resulted in a less aggressive schedule for reform, a diminished role for competition, and preservation of some characteristics (if not exact elements) of the NHS that were to be removed by the reforms.

The modifications to the reforms were met with relief in some quarters – including many clinicians – and dismay in others. Alan Milburn, who had served as Secretary of State for Health in the Labour Government, denounced the revised reform plan as insufficient, saying that the coalition's amended plans are “the biggest car crash” in NHS history, and ones that would set back the cause of reform of the NHS for many years.

Despite a number of obvious differences in aims and in the national political context, important elements of the Coalition Government reforms were similar to those that had been put forward in the US and passed into law in 2010. There were also striking similarities in some of the reactions. The experience of each nation indicates how difficult it can be to implement change in a health care system that is viewed as an important right, despite misgivings about its performance.

Both nations face an imperative to constrain health expenditure growth, though a drive to lower costs was viewed as too unpopular to pursue explicitly. Like the British Government, the US Government

“The challenges for the NHS in England have parallels in both the private and public components of the US health care system.”

faces a rising budget deficit, driven in part by rising health expenditures. In the US, rising health expenditures have an equally important effect on private sector activity. Private health expenditures account for about half of total health expenditures, and are not a component of the government budget deficit. But they affect take-home pay, costs of domestic production and the ability to raise revenues for other purposes, and therefore deeply concern employers. Thus the challenges for the NHS in England have parallels in both the private and public components of the US health care system.

In response to budget pressures and to address the distinctively American challenge of increasing the number of citizens covered by health insurance, Congress passed and the President signed into law a health reform bill that sought to improve the quality of care and improve its efficiency. In both England and the US, the goal was to promote competition and improve health care quality. And, in both nations, there were attempts to promote reorganisation of care to facilitate more integrated health care delivery.

In England, this takes the form of commissioning groups and oversight procedures intended to foster integration. The US health reform law contains diverse provisions to encourage integration. These include a variety of demonstration projects in the Medicare programme to allow groups of providers to share savings from reduced utilisation of health services; the creation of a Medicare and Medicaid Innovation Center to promote novel approaches to care delivery; and the development of accountable care organisations – new entities generally consisting of groups of physicians and hospitals.

The integrated organisations, in turn, are expected to be able to accept new forms of payment – bundled payments – that encourage integrated care. Bundled payments are more aggregated than case-rates used for hospital care under Medicare in the US and Payment by Results in England. Payments for episodes of care and for bundled procedures expose physicians and hospitals to greater financial risk and incentivise reducing costs, since generally they will not be paid more for the care of patients who need more services than average. Furthermore, the bundled payments are intended to include added payments for achieving specified levels of quality. Adapting to bundled payments with these features will require great effort and management ability; such payments impose management responsibilities on health care providers similar to those that they

“The shared assumption in the US and England is that both altered incentives and reorganised care delivery will be needed to raise the quality and efficiency of health care.”

would face under capitated payments. The required competence in management is analogous to the requirements for being a successful foundation trust or clinical commissioning group in the English NHS. The shared assumption in the US and England is that both altered incentives and reorganised care delivery will be needed to raise the quality and efficiency of health care. A key test will be whether the specific payment reforms that are adopted will lead to the desired changes in both quality and costs – an outcome that may prove elusive if flawed reforms are implemented.

Reform in both nations calls for an important role for competition within limits, and in neither nation can competition be said to be an end in itself. In the US, the creation of Health Insurance Exchanges is intended to provide individuals with enhanced access to competing private health insurance plans in well-defined markets, with extensive rules governing the features of the plans and the terms under which they would compete. Although the creation and governance of the exchanges are complex – each state may administer its exchange in a unique way, subject to broad federal oversight – in several states plans to create and maintain exchanges are already well underway. Under Medicare, accountable care organisations would compete for patients on the basis of quality and service. Thus the changes parallel the commissioning groups and role of foundation trusts in the NHS, despite significant differences in details and in the scope of competition.

The Affordable Care Act also contains provisions to create a permanent organisation – the Patient-Centered Outcomes Research Institute (PCORI) – to provide better information about the effectiveness of alternative approaches to health care and its delivery, and alternative approaches to prevention of disease. PCORI is structured as a private non-profit organisation to be funded, eventually, by a fee levied on health insurance. Its role as an agency conducting or sponsoring technology assessment led naturally to comparisons with the National Institute for Health and Clinical Excellence (NICE), the well-established and respected quango known for assessing the cost-effectiveness of health interventions for the NHS.

Great controversy over the application of cost-effectiveness analysis and over the use of comparative effectiveness information to restrict the products or services that might be covered by health insurance led

“Both the US and English reforms reflect uncertainty about service definition.”

Congress to impose a number of restrictions on the types of research to be sponsored by PCORI, and on its application in government health programmes. For example, the law prohibits Medicare from basing coverage decisions on evidence generated by PCORI. This contrasts sharply with the relationship between NICE and the NHS; NICE was created specifically to enable the NHS to make better decisions about care delivery. If NICE were not to influence NHS decisions, it would be considered a failure. The relationship between NICE and the NHS is one that many nations emulate, viewing government-funded programmes to evaluate health care as a critical element of decision-making.

Despite the interest in bundled payments, both the US and English reforms reflect uncertainty about service definition, which in turn defines what it is that different organisations compete over. Should competition be over a narrow unit of service, such as an individual laboratory test, office visit or procedure? Or should the product or service be defined in more aggregated terms? And how should the extent of the market be defined? Solely by geography, or by the range of choices that a patient might consider if he or she were not subject to any limits on the providers available?

Product or service definition varies in the NHS. GP services have been effectively bundled – in fact, as a form of partial capitation – for several years. There is competition at the service level for specialist and hospital care, made possible through the Choose and Book programme.* Competition in integrated services would be akin to selecting a complete set of care from a single integrated system, although one could imagine variants in which individuals could obtain some services outside the integrated system they enrolled in.

The fact that choice is not exercised more often under Choose and Book doesn't mean that it is unimportant. The threat of choice can nevertheless be a powerful incentive for providers. However, if referral by GP is such an important determinant of actual choices, this only works well insofar as the referring physician is knowledgeable about the consultants and is able to represent the patient's preferences and desires. The effects of a pro-competitive policy are not simply a matter of ideology; they are determined by practical considerations such as these. Meaningful choice implies that providers feel pressure to do

* www.chooseandbook.nhs.uk

better in order to attract business. That means there must be incentives for volume (of the right kind) and patients must be able to observe the relevant characteristics of providers they consider.

For advocates of change in health care financing and delivery, a striking lesson common to the recent health reform experiences in the US and England is the great influence of the political process on the form that policy ultimately takes. The Coalition Government's White Paper described a broad set of policy changes, many of which had been planned for a considerable time. Similarly, the Obama administration and Congress built upon a long-discussed set of ideas about how a health insurance expansion should be designed and how costs should be controlled. In both cases, the debate among legislators and input from the public reshaped health reform initiatives.

The Coalition Government in the UK learned that health care providers, officials in the Department of Health, and many other parties objected to features of the proposed reforms, and asked that some of the changes be implemented less hastily. In the US, many of the changes that the architects of reform originally contemplated were modified or delayed. The Independent Payment Advisory Board, for example, was created to advise Congress and propose legislation to slow Medicare expenditure growth, if growth became too rapid. But Congressional discomfort with the notion of ceding too much authority to an independent advisory board led the legislators to limit the areas in which it could make cost-saving recommendations – for example, exempting hospital payments from consideration for cost-cutting until at least 2018, and prohibiting the Board from making any recommendations that would change the amount of money that Medicare beneficiaries pay out-of-pocket. The restrictions on PCORI and the delay in the implementation of an excise tax on high-cost health plans are other examples of ways that policies were modified as the legislation worked its way through the White House and Congress.

Retrenchment by the Coalition Government following the period of public commentary reflects a similar bow to political reality in England. Specific criticisms were not the same – Americans, for example, are accustomed to market-based, competitive approaches to health care, while the British are accustomed to the use of NICE's technology appraisals to determine which innovations will be offered in the NHS. In each system, the unfamiliar elements seemed to cause

“Often advocates of reform overestimate public dissatisfaction with the status quo.”

the greatest alarm. And in both systems, concerns of health care providers influenced the way that reform was reshaped, slowing the pace as well as the degree of change.

Specialists in England, for example, successfully fought for representation on clinical commissioning groups, and in the US, hospitals ensured that the tools to limit growth in the Medicare payments they received would not be available for many years. The groups that expected to be disadvantaged by reform showed that they had formidable influence and that policy cannot be divorced from politics.

Perhaps the most important lesson from these experiences, however, is how risky it is to tamper with a costly but cherished government benefit. Often advocates of reform overestimate public dissatisfaction with the status quo, and misjudge the public’s appetite for change. Health care providers and other groups and individuals with much at stake have powerful voices. They often enjoy greater public confidence than do government officials and legislators. Health policy issues are inherently complex and experts often disagree about the solutions to policy challenges. In addition, the English and American health reform initiatives occurred at a time when public confidence in governments and professional experts was shaken by the worldwide financial crisis. For all of these reasons, the need for strong evidence that change would improve health care was greater than ever.

Concluding thoughts

“The less than complete embrace of integration may well have reflected ambivalence about mixing roles.”

The steps proposed by the Coalition Government to reform the NHS have the potential to improve system efficiency, but those looking for a dramatic break from the past will be disappointed. Furthermore, changes made to the reform plans following the public reaction will dampen and delay the changes outlined in the White Paper. That is likely to be both good and bad – some delays were almost certainly necessary, if only because they could not be implemented as quickly as the government envisioned. But for those who believe that the reforms did not go far enough, the government response to the listening exercise is a step in the wrong direction.

Competition and choice were central themes in the Coalition Government’s plans to improve the NHS. However, the original White Paper placed such strict boundaries on the scope of competition that even if the explicit retrenchment from competition announced in June 2011 was only rhetorical, there would be only a limited role for competition post-reform. If efficient care requires bundling of services, and therefore either real or virtual integration among providers, the incentives for efficiency in the reform plan are incomplete. Clinical commissioning groups, for example, do not appear to be subject to full capitation. Although they take on some of the functions of an insurer, they are not responsible for the full range of health services, and there are restrictions on their ability to integrate with secondary care providers.

The less than complete embrace of integration may well have reflected ambivalence about mixing roles, as well as anxieties about allowing organisations to become too large and powerful. Many criticised the combination of care provision and commissioning roles assigned to clinical commissioning groups. Often the criticism was based on the misapprehension that the same individuals would carry out the contracting and management functions as well as deliver patient care. In integrated organisations, of course, distinct personnel are responsible for these different functions. To enable such specialisation, the organisations need to achieve a minimum scale. And they need scale to handle other aspects of the combined role.

Successful melding of commissioning and care provision requires the ability to accept the financial risk that comes with the insurance

“Rewards for better outcomes of care are better directed towards large groups of physicians and other providers.”

function, as well as management talent, appropriate organisational structure, and well-aligned internal incentives. And rewards for better outcomes of care, as opposed to rewards for adherence to prescribed processes, are better directed towards large groups of physicians and other providers than towards individuals and small practices. The need for scale, in this instance, derives from statistical considerations, such as the need to have sufficient numbers of patients with specific conditions to be able to determine whether variation in outcomes is not solely due to chance, as well as the need to ensure that providers do not bear excessive financial risk.

Because clinical commissioning groups are not allowed to form or compete across geographic boundaries, they may not be able to reach an efficient scale, and without tight relationships with hospitals and secondary care providers they may not have sufficient scope to implement practices that would lead to efficient care. Choice can be manifested at various levels; for example, individuals could choose GPs who will provide all of their care, or they may exercise choice by selecting among different providers whenever they need a specific service or set of services, such as an elective surgical procedure. In the latter instance, Choose and Book already facilitates choice. But limitations imposed in the reform may be too stringent to expand the benefits of competition a great deal, particularly among GP practices or commissioning groups.

With significant economies of scale or scope, unfettered competition can lead to oligopoly or monopoly. That is why regulatory solutions or limits on competition are sometimes proposed. But if there are significant economies of scale, limiting scale by constraining the size of the market is a mandate for inefficiency. Anti-trust policy in England and in the US therefore seeks to balance the efficiencies that come from scale and scope against the anti-competitive behaviour that firms with large market share can exhibit. As a rule, anti-trust policy does not seek to keep firms from reaching efficient scale, but rather to regulate their behaviour so they do not exploit market power, which typically harms consumers by raising prices.

Ironically, that is not an issue in the NHS. In markets for most products and services, the main benefit of competition is lower prices. But in the NHS, there are no market-determined prices. Consumers pay nothing out-of-pocket for the care they receive, and where prices exist, such as payments to GPs and hospitals, they are regulated.

“A regulatory authority could impose limits on aggressive risk selection.”

Thus, price competition has no role. Undoubtedly this is by design, and it may serve an important political purpose. The antipathy towards price competition is so great that the government’s response to the Future Forum offers reassurances that ‘there will be new safeguards against price competition’ (DH, 2011: p5).

For a government that is seeking to control health expenditures that are rising too rapidly, this is a remarkable statement. On the one hand, the proposed policy limits scope and size, presumably because it would lead to too much market power. On the other hand, it is hard to say what harm market power exerts other than preventing full price competition. And weakened price competition is unlikely to deter a government that has publicly stated its opposition to price competition. The government’s position undoubtedly reflects, at least in part, a concern that competition may lead to lower quality as well as deliberate risk selection on the part of providers. But if these are concerns with price competition, they should be concerns without it as well, since providers already have incentives to lower their costs. There is no conflict between promoting competition and promoting the quality of care, as long as competition is over quality as well as price. Avoidance of price competition is not a strategy to improve the quality of health care. Rewarding better quality is.

By bringing market features into a governmental body, the NHS has had to meld the highly decentralised world of markets with an organisation that has had a history of significant central oversight. Despite the existence of the internal market, decisions about hiring, firing and many aspects of practice for PCTs and, to a lesser extent, hospital trusts, have long been the prerogative of NHS officials. Central control makes it possible to modify incentives in ways that are difficult in pure markets; for example, aggressive competition among commissioning groups could lead to adverse selection, which conventional risk-adjustment mechanisms might not adequately control. A regulatory authority could impose limits on aggressive risk selection. More generally, central oversight makes it possible to impose criteria that are vague and subject to rapid revision, which have the advantages and disadvantages discussed above.

At the same time, as a government body, the NHS may find it difficult to implement decisions that are necessary yet politically sensitive. Competition could well force out inefficient commissioners, primary

care physicians or hospitals, which may be politically problematic but better for patient health.

Perhaps the most important reason to encourage competition and allow flexibility in the organisation of care delivery, in commissioning and in the incentives used to reward providers, is that we simply don't know which approaches work best. A competitive, flexible approach to health care delivery will promote innovation, making it possible to learn from health care providers that adopt different approaches to delivery. A top-down, precisely prescribed approach to incentives, organisation and specific forms of care is suitable when we already know the best ways to carry out our work. If we don't know, however – and our knowledge is far from complete – flexibility offers the opportunity to learn and adapt. It will be important to give the clinical commissioning groups, and the NHS Commissioning Board, the ability to develop a wide variety of approaches to commissioning and to the organisation of care. Their experience, when accompanied by detailed evaluation of the effects of alternative approaches, can lead to public acceptance of change and enduring improvements in care.

References

- Appleby J (2011a) 'How satisfied are we with the NHS?', *BMJ* 342: d1836.
- Appleby J (2011b) 'What's happening to waiting times?', *BMJ* 342: d1235.
- Arrow KJ (1963) 'Uncertainty and the welfare economics of medical care', *American Economic Review* 53(5), 941–973.
- Audit Commission (2005) *Early Lessons from Payment by Results*. London: Audit Commission.
- Audit Commission (2008) *The Right Result? Payment by Results 2003–07*. London: Audit Commission.
- Baily MN and Garber A (1997) 'Health care productivity', *Brookings Papers on Economic Activity: Microeconomics*, 143–202.
- Balk EM, Bonis PAL, Moskowitz H, Schmid CH, Ioannidis J, Wang C and others (2002) 'Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials', *JAMA* 287(22), 2973–2982.
- Banks J, Marmot M, Oldfield Z and Smith JP (2006) 'Disease and disadvantage in the United States and in England', *JAMA* 295(17), 2037–2045.
- Bevan G and Hood C (2006) 'Have targets improved performance in the English NHS?', *BMJ* 332(7538), 419–422.
- Bloom N, Propper C, Seiler S and Van Reenen J (2010) *The Impact of Competition on Management Quality: Evidence from public hospitals*. Cambridge, MA: National Bureau of Economic Research.
- Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamotheu BK and others (2006) 'Hospital quality for acute myocardial infarction', *JAMA* 296(1), 72–78.
- Burke C (2010) *Removing the Policy Barriers to Integrated Care Developing an Integrated Approach to Clinical Commissioning Through Polysystems in Outer North East London*. London: Nuffield Trust.
- Cameron D (2011) Speech on NHS reforms, Ealing Hospital, 16 May. *New Statesman*. www.newstatesman.com/uk-politics/2011/05/nhs-health-change-care
- Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B and Roland M (2007) 'Quality of primary care in England with the introduction of pay for performance', *N Engl J Med* 357(2), 181–190.
- Campbell SM, Reeves D, Kontopantelis E, Sibbald B and Roland M (2009) 'Effects of pay for performance on the quality of primary care in England', *N Engl J Med* 361(4), 368–378.
- Chantrill C (2011) 'Pensions NHS education spending chart'. www.ukpublicspending.co.uk
- Coulam RF and Gaumer GL (1991) 'Medicare's prospective payment system: a critical appraisal', *Health care financing review, Annual supplement*, 45.
- Cromwell J, Trisolini MG, Pope GC, Mitchell JB and Greenwald LM, eds (2011) *Pay for Performance in Health Care: Methods and approaches*. Research Triangle Park, NC: RTI Press.
- Crosson B, Propper C and Perkins A (2001) 'Do doctors respond to financial incentives? UK family doctors and the CP fundholder scheme', *Journal of Public Economics* 79, 375–398.
- Davies SM, Geppert J, McClellan M, McDonald KM, Romano PS and Shojania KG (2001) *Refinement of the HCUP Quality Indicators*. Rockville MD: Agency for Healthcare Research and Quality.
- Deci EL, Koestner R and Ryan RM (1999) 'A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation', *Psychological Bulletin* 125(6), 627–668.
- Department of Health (2010) *Equity and Excellence: Liberating the NHS* White Paper, Cm 7881. London: The Stationery Office.
- Department of Health (2011) *Government Response to the Future Forum Report*. London: The Stationery Office.
- Dimick JB, Welch HG and Birkmeyer JD (2004) 'Surgical mortality as an indicator of hospital quality', *JAMA* 292(7), 847–851.
- Doran T (2011). Private communication.
- Draper D, Kahn KL, Reinisch EJ, Sherwood MJ, Carney MF, Kosecoff J and others (1990) 'Studying the effects of the DRG-based prospective payment system on quality of care', *JAMA* 264(15), 1956–1961.
- Enthoven AC (1978a) 'Consumer-choice health plan (first of two parts). Inflation and inequity in health care today: alternatives for cost control and an analysis of proposals for national health insurance', *N Engl J Med* 298(12), 650–658.
- Enthoven AC (1978b) 'Consumer-choice health plan (second of two parts). A national-health-insurance proposal based on regulated competition in the private sector', *N Engl J Med* 298(13), 709–720.
- Enthoven AC (1993) 'The history and principles of managed competition', *Health Affairs* 12 Suppl, 24–48.
- Enthoven AC (1999) *In Pursuit of an Improving National Health Service*. London: Nuffield Trust.

- Enthoven AC and van de Ven WP (2007) 'Going Dutch: managed competition health insurance in the Netherlands', *N Engl J Med* 357(24), 2421–2423.
- Enthoven AC, Crosson FJ and Shortell SM (2007) "Redefining health care": medical homes or archipelagos to navigate?, *Health Affairs* 26(5), 1366–1372.
- Feachem RG, Sekhri NK and White KL (2002) 'Getting more for their dollar: a comparison of the NHS with California's Kaiser Permanente', *BMJ* 324(7330), 135–141.
- Fisher ES and Shortell SM (2010) 'Accountable care organizations: accountable for what, to whom, and how', *JAMA* 304(15), 1715–1716.
- Fox R (2002) 'Efficiency in the NHS', *J R Soc Med* 95(12), 579.
- Gaynor M, Moreno-Serra R and Propper C (2010) *Death by Market Power: Reform, competition and patient outcomes in the National Health Service*. Cambridge, MA: National Bureau of Economic Research.
- Gneezy U and Rustichini A (2000a) 'A fine is a price', *Journal of Legal Studies*, 1–17.
- Gneezy U and Rustichini A (2000b) 'Pay enough or don't pay at all', *Quarterly Journal of Economics* 115(3), 791–810.
- Ham C (2010) *Ten Challenging Questions about the White Paper*. London: The King's Fund.
- Ham C and Smith J (2010) *Removing the Policy Barriers to Integrated Care in England*. London: Nuffield Trust.
- Hutton R and Penny T (2010) 'Britain: sparing health care from budget cuts', *Bloomberg Businessweek*.
- Iezzoni LI (1997) 'Assessing quality using administrative data', *Ann Intern Med* 127(8 Part 2), 666–674.
- Jha AK, Orav EJ, Li Z and Epstein AM (2007) 'The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures', *Health Affairs* 26(4), 1104–1110.
- Kahn KL, Keeler EB, Sherwood MJ, Rogers WH, Draper D, Bentow SS and others (1990) 'Comparing outcomes of care before and after implementation of the DRG-based prospective payment system' *JAMA* 264(15), 1984–1988.
- Kessler DP and McClellan MB (2000) 'Is hospital competition socially wasteful?', *Quarterly Journal of Economics* 115(2), 577–615.
- Laffont J and Martimort D (2001) *The Theory of Incentives: The principal-agent model*. Princeton, NJ: Princeton University.
- Lave JR (1989) 'The effect of the Medicare prospective payment system', *Annual Review of Public Health* 10(1), 141–161.
- Leavitt MO (2006) *Report to Congress: Physician group practice demonstration first evaluation report*. Washington, DC: Department of Health and Human Services.
- Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A and others (2007) 'Public reporting and pay for performance in hospital quality improvement', *N Engl J Med* 356(5), 486–496.
- Maddox GL (1999) 'General practice fundholding in the British National Health Service reform, 1991–1997: GP accounts of the dynamics of change', *J Health Polit Policy Law* 24(4), 815–834.
- Mant J and Hicks N (1995) 'Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction', *BMJ* 311(7008), 793–796.
- Maynard A and Bloor K (2010) 'Will financial incentives and penalties improve hospital care?', *BMJ* 340, doi: 10.1136/bmj.c88.
- McGlynn EA (2003) 'Selecting common measures of quality and system performance', *Medical Care* 41(1), 39–47.
- McNulty PJ (1967) 'A note on the history of perfect competition', *Journal of Political Economy* 75(4), 395–399.
- Milgrom P and Roberts J (1992) *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Moore L and O'Meara H (2009) Personal communication with the author, 8 July.
- National Centre for Social Research (NatCen) (2010) *British Social Attitudes 27th Report*.
- Newhouse JP (1993) *Free for All? Lessons from the Rand Health Insurance Experiment*. Cambridge, MA: Harvard University Press.
- NHS Institute for Innovation and Improvement (2011) Commissioning for Quality and Innovation (CQUIN) payment framework. www.institute.nhs.uk/cquin
- NHS Redbridge (2011) *About Us*. <http://redbridge.nhs.uk/temp-www4.com/aboutus/index.asp>
- Peikes D, Chen A, Schore J and Brown R (2009) 'Effects of care coordination on hospitalization, quality of care, and health expenditures among Medicare beneficiaries: 15 randomized trials', *JAMA* 301(6), 603–618.
- Petersen LA, Woodard LCD, Urech T, Daw C and Sookanan S (2006) 'Does pay-for-performance improve the quality of health care?', *Ann Intern Med* 145(4), 265–272.

-
- Pike C (2010) *An Empirical Analysis of the Effects of GP Competition*. Working Papers Series. London: Cooperation & Competition Panel.
- Porter ME and Teisberg EO (2004) 'Redefining competition in health care', *Harvard Business Review*.
- Propper C, Burgess S and Gossage D (2003) *Competition and Quality: Evidence from the NHS internal market 1991–1999*. CiteSeer.
- Propper C, Burgess S and Green K (2004) 'Does competition between hospitals improve the quality of care?: hospital death rates and the NHS internal market', *Journal of Public Economics* 88(7–8), 1247–1272.
- Rosenthal MB and Frank RG (2006) 'What is the empirical basis for paying for quality in health care?', *Medical Care Research and Review* 63(2), 135–157.
- Rosenthal MB, Fernandopulle R, Song HR and Landon B (2004) 'Paying for quality: providers' incentives for quality improvement', *Health Affairs* 23(2), 127–141.
- Rosenthal MB, Landon BE, Normand SL, Frank RG and Epstein AM (2006) 'Pay for performance in commercial HMOs', *N Engl J Med* 355(18), 1895–1902.
- Shortell SM and Casalino LP (2008) 'Health care reform requires accountable care systems', *JAMA* 300(1), 95–97.
- Shortell SM and Hughes EF (1988) 'The effects of regulation, competition, and ownership on mortality rates among hospital inpatients', *N Engl J Med* 318(17), 1100–1107.
- Sloan FA, Morrisey MA and Valvona J (1988) 'Effects of the Medicare prospective payment system on hospital cost containment: an early appraisal', *The Milbank Quarterly* 66(2), 191–220.
- Stevens S (2004) 'Reform strategies for the English NHS', *Health Affairs* 23(3), 37–44.
- Stevens S (2011) 'Is there evidence that competition in healthcare is a good thing? Yes', *BMJ* 343: d4136.
- Stigler GJ (1957) 'Perfect competition, historically contemplated', *Journal of Political Economy* 65(1), 1–17.
- Thomas JW and Hofer TP (1999) 'Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care', *Medical Care* 37(1), 83–92.
- Thompson G (2009) *NHS Expenditure in England*. London: House of Commons Library.
- Werner RM and Bradlow ET (2006) 'Relationship between Medicare's hospital compare performance measures and mortality rates', *JAMA* 296(22), 2694–2702.
- Werner RM, Kolstad JT, Stuart EA and Polsky D (2011) 'The effect of pay-for-performance in hospitals: lessons for quality improvement', *Health Affairs* 30(4), 690–698.
- Young GJ, Burgess JF Jr and White B (2007) 'Pioneering pay-for-quality: lessons from the rewarding results demonstrations', *Health Care Financ Rev* 29(1), 59–70.

For more information about the Nuffield Trust,
including details of our latest research and analysis,
please visit www.nuffieldtrust.org.uk

 Download further copies of this Viewpoint from
www.nuffieldtrust.org.uk/publications

 Subscribe to our newsletter:
www.nuffieldtrust.org.uk/newsletter

 Follow us on Twitter: [Twitter.com/NuffieldTrust](https://twitter.com/NuffieldTrust)

Nuffield Trust is an authoritative
and independent source of
evidence-based research and
policy analysis for improving
health care in the UK

59 New Cavendish Street
London W1G 7LP

Telephone: 020 7631 8450
Facsimile: 020 7631 8451
Email: info@nuffieldtrust.org.uk



www.nuffieldtrust.org.uk

Published by the Nuffield Trust.

© Nuffield Trust 2011. Not to be reproduced
without permission.

ISBN: 978-1-905030-51-4