

• NHS Reform

GETTING BACK ON TRACK

NHS REFORM

Getting back on track

Keith Palmer

King's **Fund**

© King's Fund 2006

The views expressed are those of the author and do not necessarily reflect the views of organisations with which he is affiliated.

Charity registration number: 207401

First published 2006 by the King's Fund.

All rights reserved, including the right of reproduction in whole or in part in any form.

ISBN-10: 1 85717 552 2

ISBN-13: 978 1 85717 552 3

A catalogue record for this publication is available from the British Library.

Available from:

King's Fund

11-13 Cavendish Square

London W1G 0AN

Tel: 020 7307 2591

Fax: 020 7307 2801

Email: publications@kingsfund.org.uk

www.kingsfund.org.uk/publications

Edited by Eleanor Stanley

Typeset by Andrew Haig Associates

Printed and bound in the UK by the King's Fund

Contents

List of figures and tables	v
About the author	vi
Foreword	vii
Acknowledgements	ix
Glossary	x
Summary	xv
1 Introduction and background	1
What are government policies trying to achieve?	2
The NHS reform strategy	3
Key concepts	6
Summing up	11
2 Why is the NHS in financial crisis?	13
Failure to adequately manage demand	13
High cost inflation	15
Declining hospital productivity	17
Inadequate policy levers	18
Summing up	19
3 How is the NHS likely to respond to recent policy developments?	21
2006/7 system rules	21
Payment by results	25
The commissioning framework	31
Will the government's reform objectives be realised?	35
Summing up	36
4 How can the reforms get back on track?	37
Developing strategic commissioning	37
Improving payment by results tariffs	40
Clarifying the system-regulation and management rules	42
Summing up	44
5 Conclusions	45

References	47
Appendix 1: Financial balance in the NHS	51
Appendix 2: The market forces factor	53
Appendix 3: How RAB accounting works	55

List of figures and tables

Figure 1: Emergency admissions in England by quarter	14
Figure 2: Changes in bed days and emergency admissions by strategic health authority	14
Figure 3: Use of £3.6 billion additional funding for hospital and community services in England, 2005/6	15
Figure 4: Estimated use of £4.5 billion additional funding for hospital and community services in England, 2006/7	16
Figure 5: Impact of differential ‘sunk’ capital costs on trust finances	27
Table 1: Framework for health reform in England	5
Table 2: Differences between payment by results in Australia and England	30
Table 3: Illustration of RAB accounting (1)	55
Table 4: Illustration of RAB accounting (2)	56
Table 5: Financing of income–expenditure deficits	57

About the author

Keith Palmer is a Senior Associate at the King's Fund. His other health sector involvements include acting as a non-executive director of Guy's and St Thomas' NHS Foundation Trust, a member of the South-East London Service Redesign and Sustainability Project Board, and trustee and treasurer of Cancer Research UK. He is the founder and chairman of Cambridge Economic Policy Associates, a public policy consulting firm, and for many years held senior positions in a major UK investment bank. Further details and copies of his earlier work on health reform can be accessed at www.keithpalmer.org

Foreword

To listen to some of the protagonists in the current health debate you could be forgiven for believing that the government's market-style reforms are already transforming the NHS – for either good or ill. In reality, thus far, their direct impact has been marginal; today's successes and failures are largely the result of an earlier phase of reform.

Nevertheless, there is no doubt that the new incentives and other arrangements now being put in place will have a significant effect on the way services are organised and run. Exactly what the impact will be or how it will manifest itself in different parts of the system is much less clear. This is partly because we are embarking on new territory – nothing quite like this array of policies has been attempted before – and partly because individual measures, such as practice-based commissioning and payment by results, are being introduced without necessarily a full understanding of either precisely how they will operate on the ground or, more importantly, how they will act in combination with one another.

At the same time, the pressure for the reforms to deliver is growing. The government needs to demonstrate that the extraordinary investment of the last few years has brought about higher quality, more responsive services that provide value for money. This pressure is all the greater because time is limited.

From 2008 the recent average real funding increase of more than seven per cent a year looks set to drop to around three per cent. Given that NHS deficits are already leading to service reductions in some parts of the country, and that the current configuration of too many services is neither efficient nor effective, the challenges faced are considerable.

Far reaching changes are needed to where and how care is provided, but it will be vital that they are shaped by the right forces.

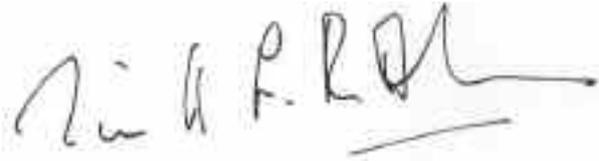
This report by Keith Palmer, one of our Senior Associates, provides a timely view of the challenges ahead and what could happen if the process is not managed effectively and perverse incentives are left to flourish.

While acknowledging that the government is committed to revising its 'operating framework' further (new guidance is expected in late 2006), the report proposes specific action to strengthen commissioning, improve payment by results, and help the whole system work more coherently.

Keith Palmer's analysis suggests that the danger is not a case of 'too much reform, too quickly' but 'too little reform, too late'. That is controversial, but what is not disputed is

that more work needs to be done to ensure the various pieces in this complex jigsaw come together to create a coherent whole.

We believe this report is an excellent contribution to what we hope will be a challenging, but ultimately constructive, debate to achieve that end.

A handwritten signature in black ink, appearing to read 'Niall Dickson', with a horizontal line underneath the name.

Niall Dickson
October 2006

Acknowledgements

I am grateful for helpful comments on earlier drafts of the paper from many individuals working in and with the NHS and the Department of Health. I would particularly like to thank John Appleby, Richard Lewis and Michelle Dixon at the King's Fund, as well as Tom McGuire, who provided many insights about the way payment by results works, and Graham Simpson, until recently the Project Director for the South-East London Service Redesign and Sustainability Project. Needless to say the views expressed in the paper are entirely the responsibility of the author.

Glossary

Capital expenditure, capital cost, capital stock, annual capital charge, cost of finance, public dividend capital (PDC), 'sunk' fixed costs

Capital expenditure refers to expenditure in a year on capital – that is, assets with a life of more than one year. Capital cost is the cost of capital assets incurred in the year. Capital stock is the total depreciated cost of capital assets incurred in all previous periods. The annual capital charge is the sum of depreciation and the cost of finance in a year. The cost of finance is the sum of interest on debt and the dividend on public dividend capital. The term public dividend capital (PDC) refers to the public sector capital invested in a trust. Hospital trusts are required to pay an annual payment to the Department of Health, called the dividend on PDC, calculated as a fixed percentage of the value of PDC provided to them. 'Sunk' fixed costs are costs incurred on capital assets in previous periods that are irreversible. The annual capital charge relating to 'sunk' fixed assets in the income–expenditure account (*see* opposite) will continue unchanged, regardless of how efficient or inefficient the hospital might be for the remaining life of the assets.

Commissioning framework

The term commissioning framework refers to the rules applying to health care commissioners, which are set out in the Department of Health publication, *Health Reform in England: Update and commissioning framework*, published in July 2006.

Demand management, referral management centre

The term demand management refers to actions taken by primary care trusts and/or GP practices to moderate the demand for health care services. Hospital demand management refers to actions taken to moderate the rate of referrals of patients to hospitals. A referral management centre assesses the clinical appropriateness of GP referrals before treatment is allowed to proceed. These centres are one of the demand-management mechanisms described in the July 2006 commissioning framework (*see* above).

Demand-side reforms, supply-side reforms, practice-based commissioning

Demand-side reforms is the name given to reforms that are intended to strengthen the ability of patients, GP practices and primary care trusts (PCTs) to shape the level, pattern and quality of services provided for NHS patients. Examples include patient choice, practice-based commissioning and PCT-contracting. Supply-side reforms aim to create more flexible and responsive providers of health care for NHS patients. Examples include

the creation of foundation trust hospitals and the use of independent sector providers of care for NHS patients. Practice-based commissioning is a form of commissioning that is carried out by GP practices. Its purpose is to provide GPs with improved incentives to reduce inappropriate hospital referrals and to increase the provision of health services closer to patients' homes.

Finished consultant episode (FCE), spell

The term finished consultant episode (FCE) refers to a measure of clinical activity performed. In the past, hospitals used this measure in health statistics to quantify their clinical activities. A patient who saw more than one consultant during a period in hospital would usually generate more than one FCE. A spell is the measure of clinical activity used in the payment by results system. Broadly, one spell measures one period spent by a patient in hospital, during which time the patient may be seen by more than one consultant. The payment by results tariff is paid for each spell of activity provided.

Income–expenditure account, income–expenditure charge, income–expenditure deficit

The NHS income–expenditure account is broadly equivalent to the profit and loss account of a private sector company. An income–expenditure charge is the accrued cost in the income–expenditure account of expenditure incurred in the year. The income–expenditure capital charge in a year is the sum of the depreciation charge, accrued interest on debt and the dividend on the public dividend capital (see opposite). In the NHS, an income–expenditure deficit is incurred if the sum of expenditure, including the annual capital charge, exceeds the accrued income in the year.

Legacy costs

The current average cost of provision for each hospital trust is a legacy of past investment and service-delivery decisions that cannot readily be reversed. The capital costs per unit of activity of 'sunk' capital at a particular trust may be well above or below the average capital cost per unit of activity for the NHS as a whole. The annual capital charge – the depreciation and cost of finance (the dividend on public dividend capital, see opposite) – must be shown as costs in the income–expenditure account for the remaining life of the assets. Whenever a trust's actual sunk fixed costs per unit of activity are higher than the national average, payment by results tariffs will not fully reimburse those costs. The unrecovered costs are referred to as legacy costs.

Market-clearing mechanism, stranded capacity, stranded costs

A market-clearing mechanism ensures that the supply of services in aggregate and in a region equals affordable demand. In competitive markets, flexible prices provide the market-clearing mechanism. If prices are fixed, there is no market-clearing mechanism. If potential supply exceeds affordable demand, there will be capacity available to provide services that remains underused because there are no purchasers able to afford to pay for more services. This available but under-used capacity is referred to as stranded capacity and the costs associated with it as stranded costs.

Payment by results (PbR), health care resource group (HRG), reference costs, reference cost index (RCI), market forces factor (MFF), scaling factor

Payment by results (PbR) is a prospective payment system where the provider is paid a fixed amount for each patient that receives a particular procedure. The fixed payment is set before the beginning of the year. The income accruing to the provider therefore varies directly with the volume of activity (measured in terms of spells, see p xi) provided. Health care resource groups (HRGs) are the categories of procedure identified in PbR. A different fixed payment is made for one spell of each HRG. Reference costs are the allocated costs of providing each HRG as reported by NHS hospital service providers. The reference cost index (RCI) is a measure of the relative reference costs of hospital service providers after making various adjustments to reported reference costs. The market forces factor (MFF) is an adjustment made to tariffs intended to take account of the non-controllable regional cost variations of hospital providers. The MFF is a scaling factor, which means that it is used to allocate a fixed sum of money between providers. As a result, if the scaling factor is inaccurate, an over-allocation to one trust must result in an under-allocation to another trust or trusts.

Productivity, efficiency factor, RPI-X

Productivity is a measure of the inputs required to produce a given volume of outputs. An improvement in productivity means that fewer resources (staff, equipment, drugs, and so on) are needed to produce a given volume of patient services. Efficiency is also a measure of the inputs required to deliver a given volume of outputs but it is measured relative to 'best practice'. The efficiency factor in payment by results (currently 2.5 per cent) specifies the amount by which productivity has to improve year-on-year if a hospital provider is to achieve income–expenditure balance, assuming that it starts the year in balance. In the regulated utility industries, prices are adjusted in accordance with a formula, RPI-X, where RPI refers to the Retail Price Index (a measure of retail price inflation) and X is the efficiency factor.

Purchaser parity adjustment (PPA)

Purchaser parity adjustment (PPA) payments refer to compensation payments made to certain primary care trusts to offset the loss of purchasing power that they would otherwise suffer with the introduction of payment by results and the resulting move from locally negotiated prices to national tariffs.

Top-slicing

Top-slicing refers to the reduction in primary care trust allocations at the beginning of the year to create financial reserves to finance historic and current NHS deficits and service improvement.

Transaction reforms

The term transaction reforms refers to reforms that are necessary to enable demand-side reforms and supply-side reforms to operate. The main transaction reform is the introduction of payment by results.

Weighted capitation formula

The weighted capitation formula is the formula used to set primary care trust funding allocations. It adjusts uniform per capita allocations, taking account of the varying levels of need across different populations, with the aim of ensuring equal access for equal need across the country.

Summary

In recent years, the NHS has seen the most sustained period of rapid funding growth ever. By 2007/8, annual spend will be 40 per cent higher in real terms than it was five years earlier. But despite the increased funding, the NHS is in deficit. In 2005/6 NHS trusts in aggregate overspent by more than £1.2 billion, and the NHS as a whole overspent by more than £500 million. More than 60 trusts incurred significant deficits, and stories of staff reductions, service cutbacks and ward closures are widespread.

So what has gone wrong?

Why is the NHS in financial crisis?

There are four interrelated factors that contribute to the explanation:

- **the failure to anticipate and manage the rapid growth in demand** for health care, especially emergency admissions
- **rapid cost inflation** – especially of pay and drug costs – which has sharply increased the unit cost of services. Pay costs were underestimated and underfunded, but trusts nevertheless had to pay them in full
- **the decline in productivity across the NHS**, resulting in increases in the cost of a given volume of services even before taking account of cost inflation
- **the policy levers** put in place to manage demand and encourage NHS providers to improve productivity were too weak. The centrally negotiated pay deals have not yet delivered higher productivity. The NHS accommodated and funded the higher level of demand at much higher average cost per patient.

In fact, the failure to achieve NHS-wide financial balance should not have been a surprise. The policy levers in place at the time were too weak to ensure that aggregate NHS expenditure remained within budget – and in 2004/5, and even more so in 2005/6, it exceeded budget. Nor were there mechanisms to direct extra money to where it would do the most good. In the event, a lot of extra money was spent on funding the rapid growth of short-stay emergency admissions.

The deficits incurred by more than 60 trusts in 2005/6 resulted in part from the differential impact of the staged introduction of the reforms, particularly payment by results (PbR), and in part from the differential costs of meeting the waiting time targets and responding to the rapid growth in emergency admissions, which was much greater in some areas than others. There were bound to be ‘winners’ and ‘losers’ emerging during the transition from

the old to the new NHS. Winners would find it easier to achieve financial balance and the losers would be more likely to incur deficits and to have to adjust over time. The financial regime that required all trusts to maintain precise income and expenditure balance in every year was a mistake – it turned a period of necessary adjustment into a perceived financial crisis.

The financial crisis was not caused by ‘too much reform, too quickly’, as some would have it. This is clear because the main instruments of reform, such as PbR, practice-based commissioning (PbC) and patient choice, only began to ‘bite’ in earnest in 2006/7. The real cause of the crisis is better described as ‘too little reform, too late’ – the result of rapid growth in funding before putting in place adequate levers to manage demand and to induce improvements in provider productivity.

How is the NHS likely to respond to recent policy developments?

Recent policy developments

In early 2006, the Department of Health introduced tough new system rules intended to restore NHS-wide financial balance. Three factors – the requirement for some trusts to budget for a surplus, top-slicing of primary care trust (PCT) allocations, and halving of the purchaser parity adjustment (PPA) – all reduce PCTs’ ability to pay for services. They will have to manage hospital demand very aggressively to keep it within affordable limits if they are to avoid incurring deficits. PCTs that are subjected to top-slicing and halving of PPA in 2006/7 will not be able to afford to fund the planned growth in hospital activity that is needed if the 18-week wait target is to be met by 2008. Increasingly, efforts by PCTs to manage demand may conflict with provider competition and patient choice.

The system rules sharply strengthen the incentives on hospital trusts to improve productivity. Hospital providers will have to reduce their unit costs on average by 4–5 per cent – and some trusts by a great deal more – if they are to maintain or restore financial balance by year end. They will no longer be able to ‘trade their way out of deficits’ because PCTs will not have the funds to pay for more activity.

If hospital trusts succeed in reducing unit costs this much, then in many hospitals there will be less need for wards, beds and some categories of staff, and some services will be downsized or closed. There will be more ‘stranded’ NHS capacity, available to provide services for patients but unable to find a PCT able to pay for them. If, on the other hand, they do not succeed in reducing unit costs sufficiently, then they will incur deficits at year end. Current indications are that there will again be a significant aggregate gross trust deficit at year end, despite major efforts within-year to reduce costs.

PbR is a vital piece of the NHS reform armoury. Unfortunately, there is clear evidence that the 2006/7 tariffs are having unintended and undesirable effects. Some trusts are likely to incur deficits despite operating efficiently because they are underfunded. Some inefficient trusts may achieve financial balance and therefore appear to be efficient only because they are overfunded. An unintended result may be to induce closure of services that appear to be making losses only because the tariffs are poorly designed and underfund the service.

In their current form, the tariffs generate strong perverse incentives on hospital trusts to maximise the amount of admitted hospital care that they provide because marginal costs are less than the tariff. It therefore makes financial balance more difficult to achieve if they support, for example, a shift of patient care out of hospital and closer to home, or management of patients with long-term conditions in ways that reduce hospital admissions. The same perverse incentives may result in hospital trusts resisting collaborative working across care networks. As a result, effective networks of care may not be developed even though the outcome would be better services for patients and lower overall costs for the NHS.

The instruments available to PCTs to manage hospital demand and allocate resources to deliver national and local service priorities are weak. This weak commissioning regime will make it very difficult for PCTs (particularly those that are top-sliced and affected by halving of PPA) to keep activity within affordable limits. It will also make it difficult for them to reshape the pattern of services and develop new services closer to home in the way envisaged in the commissioning framework.

From April 2008, when the period of rapid real spending growth ends, there will be much less extra money available to expand and improve services. Affordable activity growth is expected to be just 1–1.5 per cent per annum plus whatever additional resources can be generated from further productivity gains. So, it is important that the targeted productivity improvement in 2006/7 is achieved and sustained over the medium term by all providers of care, not just hospitals, and that commissioners have stronger levers to enable them to focus additional funding on meeting priority health care needs in future years.

Will the government's reform objectives be realised?

If the reform instruments remain as they are today, will the government's reform objectives be realised? The 2006 White Paper *Our Health, Our Care, Our Say* (Department of Health 2006f) envisages a strategic shift with fewer patients inappropriately treated in hospital and more services delivered closer to home. However, progress in bringing this about is likely to be slow for as long as PCTs have weak levers to make it happen and hospital trusts lose out financially if it does. Progress towards forging local agreements aimed at reconfiguring services to deliver care across care networks is also likely to be slow for the same reasons. Desirable service reconfiguration will also be held up until transparent, workable rules are put in place to manage and finance the restructuring process.

Other government priorities include better care for patients with long-term conditions and more spending on the prevention of illness. Progress in achieving the former will depend on changes to the tariffs to strengthen incentives on providers to minimise inappropriate hospital admissions. There is already evidence that the latter – spending on public health – is being crowded out by pressures on PCT budgets.

PCTs that have been top-sliced and seen their PPA halved in 2006/7 will struggle to avoid year-end deficits. Those that do avoid a deficit will reduce the amount of income available to their providers and thereby increase the risk that the providers will incur deficits. Achieving the 18-week wait by expanding hospital activity may prove to be unaffordable for some PCTs – especially if trusts start 2007/8 with a significant aggregate deficit and PCTs are again top-sliced. Few PCTs in regions that are in deficit will be able to afford to commission or provide new services closer to home or sustain public health programmes for as long as they are required to pay for almost all hospital services at full tariff.

There will continue to be many trusts in deficit and some of the deficits will be large. The distribution and magnitude of hospital trust deficits will not be primarily a function of their relative efficiency but will depend on, among other things, the differential impact of the unintended and undesirable effects of the tariffs, the differential impact on PCT purchasing power of top-slicing and halving of the PPA and whether the trust started the year with a deficit. There is a risk that the quality of patient care at hospital trusts with large deficits will suffer as they seek urgently to restore financial balance.

Additional independent sector capacity will increase choice options for patients, but will also increase stranded NHS capacity for certain services. Very few new private finance initiative (PFI) hospitals will be required, as the productivity of NHS providers increases and the demand for hospital services flattens off. The inflexibility of existing PFI contracts will constrain service reconfigurations and tend to bias solutions in favour of expanding activity at PFI hospital sites.

By the end of 2006/7, overall NHS financial balance will be restored if the aggregate recurrent and legacy deficit of trusts is less than the value of reserves created by top-slicing and halving of the PPA. Current indications are that the reserves will not be sufficient and that overall balance may be achieved only by cutting public health, training and education spending. In future years, continuing top-slicing of PCT allocations is likely to be necessary if NHS financial balance is to be assured. This is because there will be a continuing need for sizeable reserves to fund unplanned growth in hospital demand paid at full tariff, should it occur, as well as to finance any remaining recurrent and legacy deficits and future restructuring costs.

Top-slicing of PCT allocations and halving of the PPA risk turning equitable gross PCT allocations into inequitable net allocations. Health inequalities in some PCT areas may increase as a result. Because all hospital activity covered by tariffs must be paid for at the full tariff, primary care and mental health services may suffer at the expense of the acute sector as PCTs cut spending to restore financial balance.

In summary, the new system rules, the 2006/7 tariffs and the new commissioning framework do address some of the causes of the financial crisis. They should help restore overall NHS financial balance, strengthen the ability of PCTs to manage the cost of higher-than-planned emergency admissions and strengthen incentives on NHS providers to reduce unit costs. However, the combination of weak commissioning instruments and perverse incentives embedded in the tariffs are likely to result in slow progress towards achieving other key government objectives, such as better care for patients with long-term conditions, more care closer to home, sustained improvements in the quality of hospital care and more focus on the prevention of illness.

How can the reforms get back on track?

There are three key areas where improvements to the reform instruments are needed if the government's reform objectives are to be more fully achieved:

- **development of a stronger strategic commissioning regime** that will enable commissioners more effectively to shape the structure of supply to meet national and local health care priorities in the way envisaged in the recently published commissioning framework. This will require that they are given much greater influence over the share of their budgets that are spent on elective hospital services

- **improvements to the design of PbR tariffs** to eliminate the unintended and undesired effects referred to above and to strengthen incentives on hospital trusts to support better care for patients with long-term conditions, more care closer to home and greater collaboration across care networks
- **clarification of the system rules** that will apply to regulate and manage the ‘new’ NHS market and manage and finance desired service reconfigurations.

Specific suggestions for a strengthened commissioning regime and better design of PbR tariffs are set out and discussed in Section 4, *see pp 37–42*.

1

Introduction and background

In the early years of this decade, there was tremendous optimism around the NHS. The government had just committed to huge increases in funding to make the NHS fit for the 21st century. The reform programme set out in *The NHS Plan* (Department of Health 2000) and subsequently in *Delivering the NHS Plan* (Department of Health 2002) provided a convincing analysis of why change was needed, and how it could be achieved.

Since 2000, NHS funding in England has grown by about 7 per cent per annum in real terms. By 2007/8, annual spend will be 40 per cent higher in real terms than it was five years earlier, and will account for about 8 per cent of GDP. In 2005, the King's Fund reported 'significant improvements in most areas that the government had focused policies on' but 'the NHS as a whole has not yet been transformed' (King's Fund 2005).

However, despite the rapid growth of funding, in 2005/6 NHS trusts in aggregate overspent by more than £1.2 billion, and the NHS as a whole overspent by more than £500 million (Department of Health 2006a; Appleby 2006a; King's Fund 2006c). More than 60 trusts incurred significant deficits, and turnaround teams were sent in to find out why. Although until recently a key argument for higher NHS spending was a shortage of capacity in the NHS, the press is now full of stories about staff reductions and ward closures. The NHS Confederation warns that the next two years will be 'the most challenging we have ever had to face' (Health Service Journal 2006a). The NHS is widely regarded as being in financial crisis. There is a widespread sense that system transformation is on hold while trusts struggle to maintain or restore short-term financial balance.

So, how did we reach this position, and what can be done about it? These are the questions that this paper sets out to answer, focusing on three areas.

- **Why is the NHS in financial crisis?** Why is the NHS in deficit six years into its most sustained period of rapid growth in funding ever?
- **How is the NHS likely to respond to recent policy developments?** With new system rules, wider adoption of payment by results (PbR) and the new commissioning framework in place, can we anticipate that there will now be more rapid progress towards achieving the government's wider health reform objectives?
- **How can the reforms get back on track?** What further improvements to the design of the reform instruments would improve the prospects for more fully achieving the government's key health care objectives?

What are government policies trying to achieve?

Any analysis of health care reforms must start with a clear picture of what government policies are trying to achieve. What are the broad objectives, and what are the specific targets? If these are clear, then we can test the design of the reform instruments by asking whether they can be expected to change behaviour and allocate resources in ways likely to achieve the targets and deliver over time the overall objectives.

At a high level, the overarching objective is clearly to provide a high-quality, patient-responsive health care service that is available to all and (largely) free at the point of use. The targets set by government in *The NHS Plan* (Department of Health 2000) focused heavily on increasing hospital capacity, with the aim of reducing hospital waiting times. Key targets included:

- a maximum six-month inpatient wait for elective care
- a maximum three-month wait for first outpatient appointments
- a reduction in the maximum accident and emergency (A&E) wait to four hours
- 31- and 62-day waits for cancer treatment.

By 2005/6 these initial targets had largely been met, leaving the much more demanding 18-week wait target to be achieved by 2008.

In recent years, the government's priorities for health care have evolved quite markedly. The *NHS Improvement Plan* (Department of Health 2004) set out a vision for health care that is much less hospital-centric and focuses more on improving care closer to home. The more recent White Paper *Our Health, Our Care, Our Say* (Department of Health 2006f) reinforces this shift in emphasis, describing a 'new strategic direction' with 'more services... closer to people's homes' (p 20), with more choice for patients in primary care, community services and social care, and a greater emphasis on the prevention of disease and putting people more in control of their own health and care.

As part of this shift of emphasis, the 2006 White Paper envisages the creation of new types of non-hospital services located in the community. This implies a significant shift of resources, from hospitals into the community. Section 3 of this paper (see pp 23–5) addresses the question of whether the reform instruments that were designed to induce growth in hospital activity have yet been adapted to serve this new 'care closer to home' agenda.

Recent work by the National Leadership Network (2006) argues that district general hospitals need to reinvent themselves. The network says that in future local hospitals will remain important, but that hospital providers will increasingly have to work in collaboration with other providers of care, operating as part of multi-hospital networks of care. In the network's view, this will provide patients with more appropriate, more cost-effective services. This concept has been widely supported by the Department of Health, the NHS Confederation and senior hospital clinicians. It has important implications for the future provision of hospital and community care, and will require significant reconfiguration of hospital services across the country. Another question addressed in Section 3 (see pp 23–5) is whether the reform instruments will support and enable this sort of desirable services reconfiguration.

The system rules for 2006/7 state that the specific short-term priorities are ‘substantially to reduce mortality rates from heart disease and stroke, to support people with long-term conditions, to reduce overall emergency bed days by 5 per cent and to reduce the maximum waiting time for hospital treatment to 18 weeks’ (Department of Health 2006h). The reforms ‘will support the development of high-quality services by embedding the right balance of incentives, transparency, plurality of providers and patient choice. Better commissioning... will be critical’ and the aim of the reforms is to change the health system ‘so that change is driven more by incentives to respond to patients than by top-down targets’ (Department of Health 2005a). Section 3 (see pp 35–6) considers whether these priorities are likely to be achieved given the current design of the reform instruments.

The NHS reform strategy

We have seen what government policies are trying to achieve. We now examine the reform instruments that have been devised by government with the aim of delivering these objectives.

At the heart of the reforms is the notion that the NHS should change from being a largely provider-led organisation to one that is more patient led and patient responsive. In late 2005 the Department of Health published what it termed its ‘framework for health reform in England’ (Department of Health 2005a), which set out how that is to be achieved. It explains that there are four components to the framework:

- demand-side reforms
- supply-side reforms
- transaction reforms
- system-management reforms.

The Department of Health’s description of the reforms and what they are intended to achieve is summarised below and in Table 1 (see p 5).

Demand-side reforms

Demand-side reforms focus on more choice and stronger ‘voice’ for patients. The reforms are as follows.

- **The weighted capitation formula** used to set primary care trust (PCT) allocations is intended to ensure equitable funding of health care across the country, and hence to help ensure equal access for equal need.
- **The commissioning framework** is intended to provide PCTs with the instruments to shape the pattern of services to deliver national and local health care priorities.
- **Practice-based commissioning (PbC)** aims to provide GPs with incentives to reduce inappropriate hospital referrals, and to induce greater provision of cheaper services closer to patients’ homes.
- **Patient choice** is intended to empower patients, to put pressure on hospital providers to improve the quality of elective care, thereby attracting more patients and/or avoiding losing them to other providers, and to induce provision of services that better match patients’ needs.

- **Patient 'voice'** is intended to ensure that commissioners are required to be more responsive to the needs and preferences of patients and their local communities.

Supply-side reforms

Supply-side reforms aim to create more flexible and responsive service providers. There are two key supply-side initiatives, as follows.

- **Foundation trusts** are autonomous entities operating within the NHS. They have greater operational and financial freedoms than NHS trusts. They are not subject to Department of Health direction, nor are they performance managed by strategic health authorities (SHAs). Instead, they are regulated by Monitor, the statutory body created for this purpose, and by the Healthcare Commission. Foundation trusts are required to work in partnership with other trusts in the NHS. The stated intention is that foundation trusts will be more patient responsive and innovative NHS hospital providers.
- **Independent sector providers** are private sector companies that provide contracted clinical services for NHS patients paid for by the Department of Health or NHS commissioners. First-wave contracts were for the provision of elective hospital diagnostic and treatment services. Second-wave contracts extend independent sector provision into community diagnostic and treatment services. This initiative is intended to increase capacity, stimulate greater supply-side innovation and extend patient choice.

Transaction reforms

In order for patients to be offered a plurality of providers under the policy of patient choice, there is a need to create a payment mechanism by which money follows the patient. In addition, there is a need for commissioners and providers to have access to better information on which to base their supply and demand decisions. There are two transaction reforms aimed at meeting these needs.

- **Payment by results** is a prospective payment arrangement in which the provider is paid a fixed amount for each patient that receives a particular procedure. The fixed payment for each procedure is based on the estimated cost of providing the procedure, averaged across all providers in England. This uniform price is then adjusted by the market forces factor (MFF) to derive trust-specific tariffs. The MFF-adjusted tariffs paid to providers for the same procedure vary considerably across trusts. In 2006/7, PbR tariffs apply to about two-thirds of all hospital activity.
- **Generation of activity and activity-based cost information** is an essential ingredient of a system in which commissioners have to manage demand and allocate limited resources to priority uses and in which providers have to decide whether or not to seek to offer more or less services to patients.

System-management reforms

The Department of Health recognises that there is a need for a transparent rules-based framework for two distinct but related areas:

■ **regulation of the new NHS 'market'** to support safety, quality, fairness and equity will apply to all health care commissioners and providers

■ **a new performance management regime** to ensure value for money and overall NHS financial balance will apply to all publicly funded commissioners and providers.

The Department of Health was expected to publish the proposed system-management reforms earlier in 2006 but these are now not expected until late 2006.

Table 1, taken from the Department of Health's framework document, summarises the way the government views the initiatives, aims and reform instruments fitting together.

TABLE 1: FRAMEWORK FOR HEALTH REFORM IN ENGLAND

	Demand-side reforms	Supply-side reforms	Transaction reforms	System-management reforms
Initiatives	<ul style="list-style-type: none"> ■ More choice ■ Much stronger voice for patients 	<ul style="list-style-type: none"> ■ More diverse providers ■ More freedom to innovate and improve services 	<ul style="list-style-type: none"> ■ Money follows patients ■ Rewarding best and most efficient providers ■ Giving others incentive to improve 	<ul style="list-style-type: none"> ■ System-management, regulation and decision-making to support safety, quality, fairness, equity and value for money
Aim	<ul style="list-style-type: none"> ■ To create more knowledgeable, assertive and influential users of services 	<ul style="list-style-type: none"> ■ To create more flexible, responsive and innovative service providers 	<ul style="list-style-type: none"> ■ To ensure impact of patient choice is understood and that good provider response is rewarded 	<ul style="list-style-type: none"> ■ To ensure safety and safeguard core standards in all services ■ To provide transparent rules-based framework for key management and decision-making functions in a more dynamic system
Policies	<ul style="list-style-type: none"> ■ Patient choice ■ Commissioning framework including practice-based commissioning ■ Commissioning a patient-led NHS ■ Information for patients ■ Public and patient involvement 	<ul style="list-style-type: none"> ■ NHS foundation trust status ■ Wider range of providers including private and voluntary sector providers ■ Workforce reform 	<ul style="list-style-type: none"> ■ Payment by results ■ Management information 	<ul style="list-style-type: none"> ■ Defining new performance management and regulatory functions ■ Processes for ensuring quality, licensing providers and price setting, competition policy and performance and support regime ■ Wider review of regulation

Source: Adapted from Department of Health 2005a, p 17

The constraints on hospital trusts

The way in which hospital trusts respond to the incentives created by the reforms described above is strongly influenced by the requirement on them to achieve the ‘must do’ financial targets (Palmer 2002). The key financial target that has to be met in each year is the achievement of income–expenditure balance. The income–expenditure balance target requires trusts to ensure that total income equals total expenditure in each year. (In 2006/7, the rules were tightened for many trusts, requiring them to budget for a 1 per cent surplus at year end.) Total expenditure includes two elements: revenue expenditure such as pay and drug costs; and the two components of the capital charge, namely, depreciation and the financing charge (in the NHS called the dividend on public dividend capital (PDC)). Hospital trusts have little flexibility to respond to unplanned income or cost shocks. They are not permitted to retain reserves and the dividend on PDC has to be paid in full each year. They can only respond to emerging unplanned deficits by reducing their controllable costs before the year end.

The board of a NHS trust and its chief executive can be fired by the Department of Health or the SHA if they fail to achieve the financial targets – and many have been. The severity of the punishment for failing to meet them has caused many trust boards to give absolute priority to achieving the ‘must do’ targets, with important consequences for the way they respond to the incentives created by the reform instruments.

Having looked at what government policy is trying to achieve, and the nature of the NHS reform instruments, we will now set out the key concepts and relationships that will enable us to understand the implications of the reforms.

Key concepts

In order to get an insight into how the reform instruments are operating in practice, we must first understand a number of key concepts and health-system relationships. This section addresses three overarching questions.

- What is the meaning of service affordability and why is it important?
- Why do PbR tariffs ‘force’ productivity improvement?
- How does the reformed NHS ‘quasi-market’ work?

The answers to these questions help to explain why the NHS is in financial crisis and how commissioners and providers are likely to respond to recent policy initiatives. (For a more in-depth analysis of the key concepts and health-system relationships, see Palmer 2005a.)

What is the meaning of service affordability and why is it important?

If PCTs are to be individually and in aggregate in balance in any year, then their expenditure on services must equal the money that is available to them. Their expenditure consists of the sum of purchases from primary care (the quantity of primary care services they purchase times their average cost per unit) and purchases from hospitals (the quantity of hospital services they purchase for patients, multiplied by the average tariff for those services). From this simple relationship flow several important implications.

First, if the PCT budget in a year is fixed (as it is) and the prices are fixed (as most are in PbR) and the share of hospital activity in total activity remains unchanged, then the total volume of services that PCTs can afford to purchase is also fixed. Three things follow.

- **No guarantee that all hospital activity provided will be affordable for PCTs** If hospitals provide more services for patients than are affordable given PCTs' fixed budgets, then the PCTs will not be able to afford to pay for them and they will incur deficits. Since tariffs are fixed, PCTs can only avoid deficits by ensuring that the volume of services that are provided does not exceed what is affordable.
- **No guarantee that all hospital capacity will be used** There can be no assurance that the amount of hospital capacity that is available to supply patient services will equal the volume that PCTs can afford. If the potential supply of patient services is greater than the volume that PCTs can afford, and PCTs succeed in limiting demand to affordable levels, then this will result in 'stranded capacity', meaning NHS capacity that is available to provide patient services but that remains under-used because there are no PCTs able to afford to pay for the extra services.
- **Reducing PCT allocations directly reduces the volume of affordable services** If part of PCTs' budgets are pre-empted, for example, by 'top-slicing' their allocations or requiring them to purchase minimum volumes of services from certain providers (such as independent sector providers), then the volume of services that they will be able to afford to purchase from other providers will be directly reduced. Increasing capacity without a corresponding increase in PCT purchasing power will tend to increase stranded capacity.

Second, if the budget is fixed and the share of hospital activity in total activity remains unchanged, then the volume of patient services that PCTs can afford to purchase can only be increased if the average cost of providing services (and therefore tariffs in future periods) is reduced. This highlights the crucial importance of improving the productivity of providers, thereby reducing unit costs and increasing the volume of patient services that can be provided out of fixed PCTs' budgets.

Another way of looking at the importance of productivity improvement is to ask how many extra patient services PCTs can afford to purchase year on year, given the announced increases in their funding allocations. Starting from balance, and assuming no change in the shares of hospital and primary care, the following is true:

$$\text{growth in affordable activity} = \text{growth in funding} - \text{cost inflation} + \text{productivity gains}$$

This says that the affordable growth in patient services for any given growth in funding is greater if cost inflation is lower and productivity growth is higher. If PCTs started the year with a deficit, then affordable activity growth in the following year would be lower than this because they must also eliminate the recurrent deficit and repay past deficits out of current allocations. (For more information about affordability and productivity improvement, see Appendix 1, pp 51–2.) We use this relationship later in the paper to highlight the importance for service improvement of sustained productivity improvement in the NHS.

Third, if PCTs are to increase spending on care closer to home, they will have to reduce the volume of services that they purchase from hospitals because they cannot reduce the prices. As a result, demand for hospital services will go down, and the amount of stranded hospital capacity will tend to rise. The amount of additional non-hospital activity that PCTs can afford to purchase will depend on how much they are able to reduce the volume of hospital services that they purchase, and on whether care closer to home is more or less expensive per patient than hospital care. The lower the cost per patient of care closer to home, the more additional care they will be able to afford for any given reduction in hospital activity.

Why do PbR tariffs ‘force’ productivity improvement?

PbR requires PCTs to pay providers the fixed tariff price for each unit of service they provide – in other words, for each patient that receives a particular procedure. If funding in the tariffs is reduced by 2.5 per cent, then the provider is ‘forced’ to reduce its average costs (in other words, to improve its productivity) by 2.5 per cent and maintain the volume of services that it provides, if financial balance is to be achieved. Reducing the volume of activity is not an option because income would reduce if the number of patients treated fell.

In contrast, a provider not subject to PbR could respond to a 2.5 per cent reduction in funding either by maintaining the volume of services provided and reducing average costs by 2.5 per cent or by leaving average costs unchanged and reducing the volume of services by 2.5 per cent. It is because PbR pays only for outputs actually provided that hospital trusts can be ‘forced’ to improve productivity. This is the reason that the efficiency factor in the tariffs is the key driver for productivity improvement across providers subject to PbR.

No similar mechanism exists to ‘force’ productivity improvement by providers not currently subject to PbR. Consequently, whereas hospital providers are forced to reduce unit costs but not the volume of services provided for patients, providers not subject to PbR and subject to strong budget pressures may end up cutting patient services rather than cutting their unit costs.

How does the reformed NHS ‘quasi-market’ work?

The way in which the commissioning (purchasing) of services by PCTs has been separated from the supply of services by NHS and other providers, with fixed tariffs paid for hospital services, creates what has been termed a ‘quasi-market’ for health care services (Le Grand 1993, 2003). However, it is a quasi-market with a number of very unusual characteristics, which are described below.

- **Unlimited potential patient demand** Potential patient demand is effectively unlimited because health care in the NHS is largely free at the point of use. The public is increasingly well-informed and assertive. If members of the public were free to choose how much, when, where and what type of services they wanted (and which drugs they wanted prescribed), they could impose unaffordable costs on PCTs.

Commissioning is therefore partly a process that seeks to intermediate between potential demand and affordability, seeking to obtain the greatest possible patient benefits given limited available resources. PbC, patient choice and PCT contracting all

play important roles in determining the affordable level and pattern of services provided for patients. In the past, lengthening waiting lists and waiting times were used to ensure that activity remained affordable. With the policy of reducing maximum permissible waiting times, this is no longer an acceptable way to manage demand.

- **Demand limited by affordability** As noted above, if PCTs fail to limit the volume of hospital services provided to what is affordable, they will incur deficits. As a result, whether they do in fact incur deficits depends largely on whether they have available to them effective instruments to control the volume of referrals to hospitals. Small volumes in excess of what is affordable can generate sizeable PCT deficits because they are required to pay the full average cost-based tariff for every unit of services provided.
- **Difficulties in managing hospital demand** It is very difficult for PCTs to manage non-elective demand because hospital providers have a duty to provide immediate care for all non-elective patients that turn up in A&E and PCTs have few effective levers to manage the rate of attendances and admissions. PCTs also have weak levers to manage the rate of GP and consultant-to-consultant hospital referrals because these are determined by GP practices and hospital consultants.
- **Incentives for hospitals to increase admissions** Hospital providers are entitled to be paid for almost all elective services that are funded under PbR at fixed tariffs. The tariffs are based on the national average cost of providing the relevant procedure. This means that the marginal revenue – that is, the extra income received by the provider for treating one additional patient or the income lost as a result of treating one patient less – equals the average cost of providing that service.

The costs of a hospital provider are made up of three types of costs:

- **variable costs**, which vary directly according to the number of patients treated (for example, drugs)
- **semi-fixed costs**, which are fixed in the short term but can be varied over a number of years (for example, staff costs)
- **fixed costs**, which cannot be varied over a number of years even if the volume of activity changes a lot (for example, land and buildings).

The South-East London Service Redesign and Sustainability Project team (unpublished author analysis) estimates that the typical hospital cost structure is about 10–15 per cent variable costs, about 60–75 per cent semi-fixed costs and about 15–25 per cent fixed costs.

In the short term, for small changes in volume, variable costs of all hospital providers are much less than average costs. Because marginal revenue is greater than variable costs, hospital providers are subject to strong incentives to increase activity and avoid any reduction in activity. This is because more activity will increase income by much more than it increases costs in the short term, and so will help achieve or restore annual financial balance. Conversely, less activity will reduce income by much more than costs can be reduced in the short term, making it much harder to achieve or restore annual financial balance. Moreover, hospitals will seek to maximise admitted hospital care because this is where the extra financial benefit (the difference between marginal revenue and variable cost) is greatest.

- **Lack of market-clearing mechanism** In a true market, the balance between supply and demand is achieved through the signals conveyed in flexible prices. In the NHS, because prices are fixed, there is no mechanism to ensure that the supply of services (in aggregate or in a region) equals the affordable demand for those services. If demand for services exceeds what is affordable, then one of three things must happen:
 - the demand is suppressed by PCTs (through such mechanisms as referral management centres) to ensure expenditure does not exceed what is affordable, in which case hospitals may have stranded capacity
 - activity exceeds what is affordable and PCTs agree to pay hospitals for all services provided, in which case the PCTs will incur deficits
 - activity exceeds what is affordable and PCTs refuse to pay for unaffordable activity in which case there will be a dispute between the commissioner and the provider. (Disputes of precisely this nature currently exist between a number of PCTs and foundation trusts (McIntosh 2006).)

In all three cases, unplanned income–expenditure deficits are likely to be incurred either by the PCT or by the provider.

- **Legacy costs** In any transition from a system in which services are purchased at local prices to one in which uniform national tariffs are now paid, there will always be legacy costs. These are costs incurred in the past that would not have been incurred had the reforms been in place at the time because, given the new rules, the income derived from the services is insufficient to cover the costs – even if the provider operates efficiently. Controllable legacy costs, such as too many beds or too many employees, can be managed away over a number of years but in the short term, transitional income–expenditure imbalances are bound to occur. Non-controllable legacy costs such as expensive ‘sunk’ fixed costs (for example, on land and buildings or private finance initiative (PFI) contract payments) often cannot be managed away, and trusts with these legacy costs will tend to incur persistent deficits for many years, even if they are more efficient than the average trust.
- **Disincentive for hospital trusts to support care-closer-to-home agenda** Reductions in admitted hospital activity make income–expenditure financial balance harder to achieve for all hospital providers. As a result, it is not in their interest to actively support PCT efforts to shift patient care out of hospitals and closer to home, as proposed, for example, in the 2006 White Paper, *Our Health, Our Care, Our Say* (Department of Health 2006f). Nor is it in their financial interest to actively support case management of patients with long-term conditions, as the aim of this is to keep patients out of hospital as much as possible. Of course, many clinicians support the principle of minimising inappropriate hospital admissions, but their trusts have to balance this goal against the immediate priority of achieving financial balance.
- **Disincentive for hospital trusts to support collaborative working across care networks** Hospital providers are also unlikely to support proposals to work in collaboration with others as part of multi-hospital networks of care if this results in reductions in their own activity and income, and therefore makes financial balance more difficult to achieve in the short term. Most reconfigurations to create networks of care providers do create losers as well as winners, even though the overall result is often better care for patients and lower costs for the NHS.

Summing up

So far we have looked at the government's objectives for the health reform programme and the reform instruments that they are deploying to bring about the desired changes. We have also set out the concepts and relationships that enable us to understand how the reform instruments are impacting on commissioners and providers. Equipped with this understanding, we turn now to the first of the questions posed in this section (*see p 1*) – why is the NHS in financial crisis?

2

Why is the NHS in financial crisis?

There have been significant improvements in patient care over the past five years. Maximum waiting times have been sharply reduced. Patient satisfaction surveys show that a high and improving proportion of patients are satisfied with the quality of their care. While the Healthcare Commission reviews of standards and performance note many continuing shortcomings, they do provide independent endorsement of government claims of steady improvement in the quality of care in most trusts (Healthcare Commission 2005, 2006). Hospital-acquired infections are now being tackled vigorously, with significant improvements across most of the NHS.

However, these important gains have been achieved at enormous financial cost. Despite massive funding increases, there is a widely shared perception of financial crisis in the NHS. How did we get to this position? There are four major causes:

- failure to adequately manage demand (especially emergency admissions)
- high cost inflation (especially pay and drug costs)
- declining hospital productivity
- inadequate policy levers.

This section addresses each of these causes in turn.

Failure to adequately manage demand

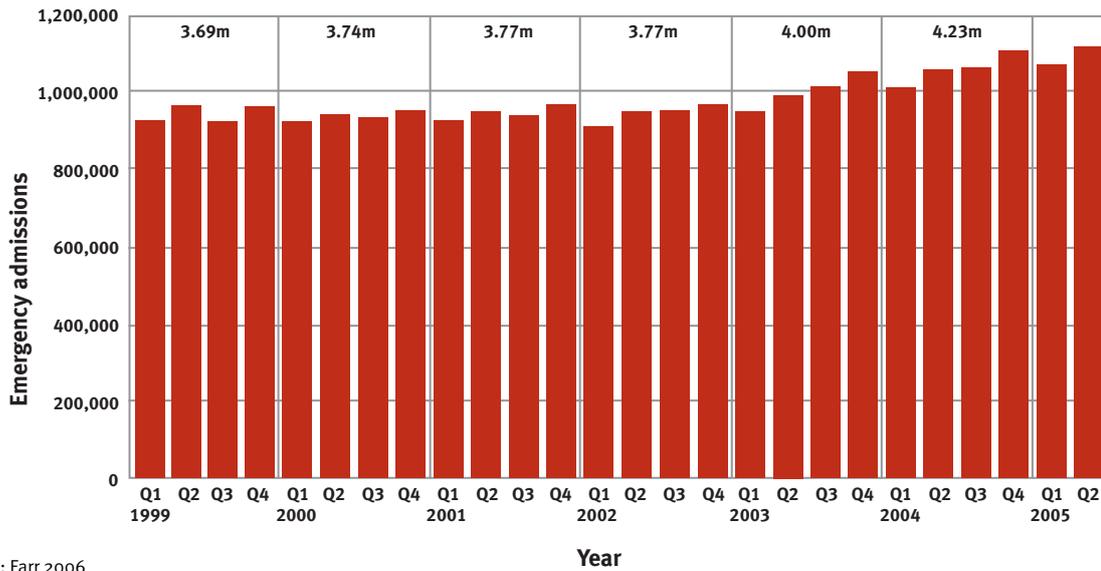
Reducing maximum waiting times for patients is not simply a matter of clearing a waiting-list backlog (Harrison 2000). As maximum waiting times in accident and emergency (A&E) are reduced, the demand for emergency patient care may increase. Similarly, the demand for elective patient care (for example, for hip operations or digital hearing aids) may increase as available funding increases, simply because GPs slow the rate of additions to waiting lists when waiting times are long and may increase them again when waiting times are shorter. More generally, the concept of supply-induced demand, where the amount of activity rises when funding increases and/or new services open, is widely understood.

Over the past three years, there has been particularly rapid growth in emergency attendances at A&E departments, and in emergency hospital admissions. Figure 1 (*see overleaf*) shows that emergency admissions grew on average by about 6 per cent per annum over the period 2003/4 across the NHS in England, and rapid growth has continued in 2005. Figure 2 (*see overleaf*) shows that there was considerable variation around this average, with some regions (including much of London) witnessing annual growth rates of emergency admissions of more than 10 per cent.

One reason for the rapid growth in A&E attendances may be that the publicity surrounding the successful reduction in A&E waiting times caused some patients who could have

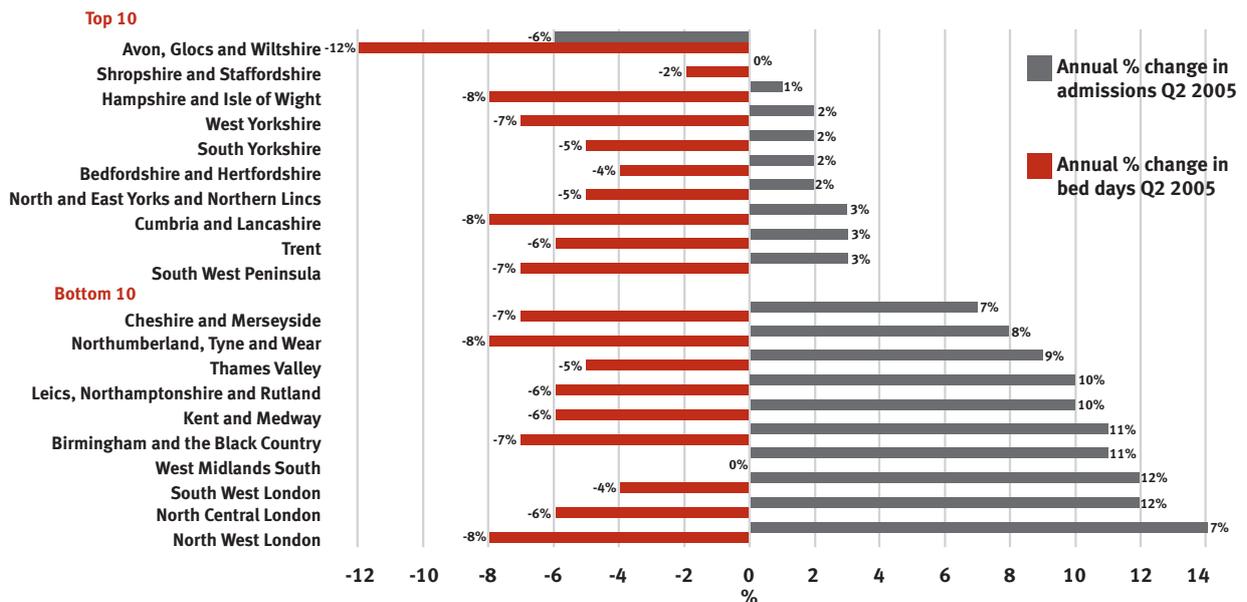
visited a GP to go to A&E instead because they could get a quicker service. In London, sample surveys of patients attending A&E indicate that 40–60 per cent of all attendees could more appropriately be seen by a GP (unpublished figures from South-East London Service Redesign and Sustainability Project).

1 EMERGENCY ADMISSIONS IN ENGLAND BY QUARTER



Source: Farr 2006

2 CHANGES IN BED DAYS AND EMERGENCY ADMISSIONS BY STRATEGIC HEALTH AUTHORITY



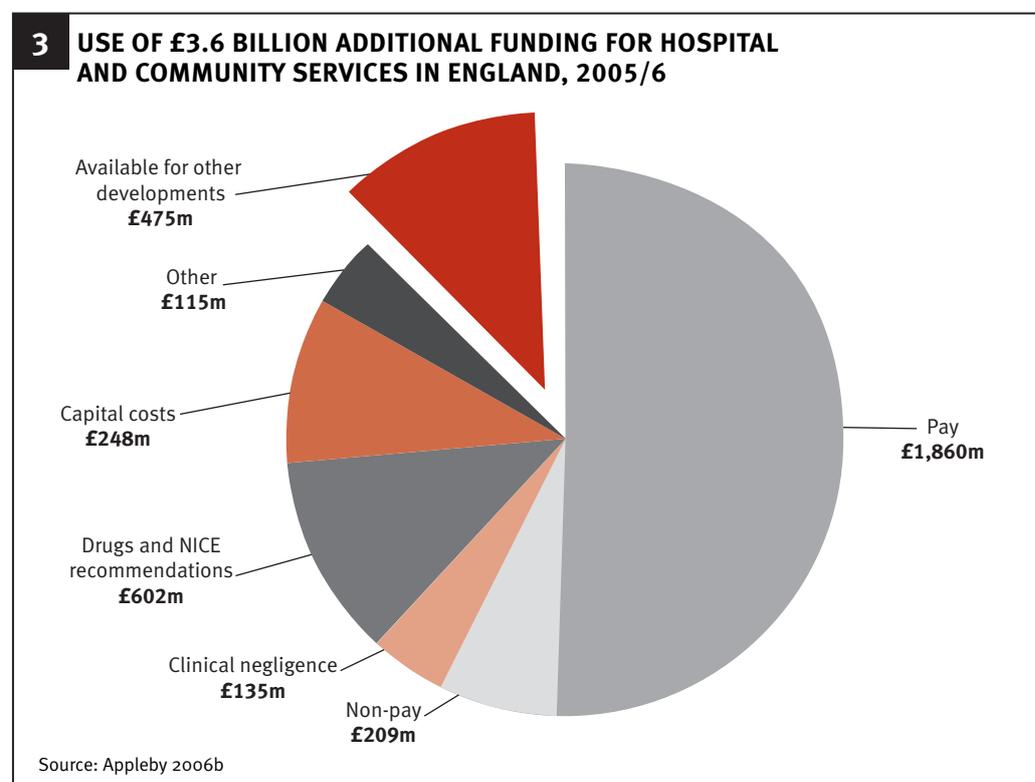
Source: Farr 2006

The rapid growth in A&E attendances converted into rapid growth in emergency admissions. However, at the same time, as Figure 2 shows, there was a reduction in the number of emergency bed days, suggesting that many of the extra admissions were for very short stays. Why this rapid growth in emergency admissions took place, and whether all admissions were clinically appropriate, is not known. However, it may be relevant that A&E departments were under great pressure to meet the maximum four-hour wait, and overnight admission for observation and diagnostic tests was one (expensive) way of meeting the target. It may also be relevant that each emergency admission brought with it more extra income than the extra cost that is typically involved in an extra overnight admission.

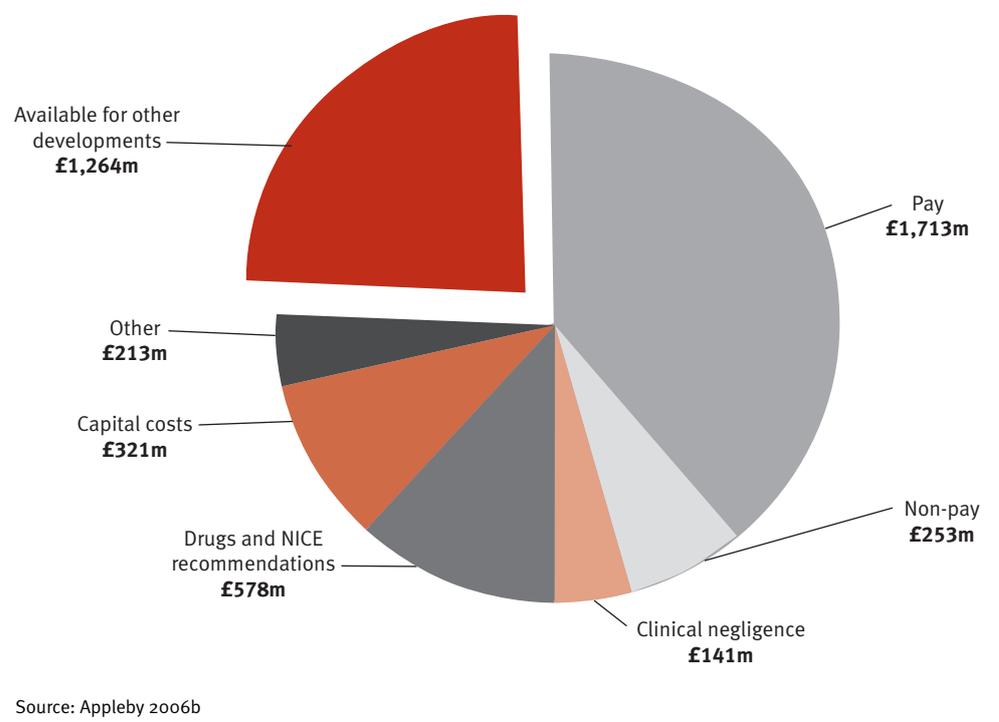
Whatever the reasons, it is clear that a lot of extra money was spent by primary care trusts (PCTs) in funding the rapid growth of emergency attendances and admissions.

High cost inflation

Since 2002, there has been a high rate of cost inflation in the NHS. In particular, average pay costs per employee have risen sharply as a result of a series of nationally negotiated pay awards. In 2005/6, 87 per cent of the extra funding for hospital and community health services in England was absorbed by higher pay costs (52 per cent), increased drug costs (17 per cent) and other unavoidable cost pressures – leaving only £475 million of uncommitted funding to pay for more activity and service improvements (see Figure 3, below). This was much less than was needed to fund the growth of non-elective and elective demand. The Department of Health has estimated that cost inflation in 2006/7 will be 6.5 per cent. The King’s Fund has estimated that 72 per cent of the extra funding in 2006/7 will be absorbed by higher pay and other cost pressures (see Figure 4, overleaf).



4 ESTIMATED USE OF £4.5 BILLION ADDITIONAL FUNDING FOR HOSPITAL AND COMMUNITY SERVICES IN ENGLAND, 2006/7



Pay cost pressures were particularly noteworthy. The new GP contract increased average GP earnings by 25 per cent over two years (BBC News online 2006a). Although the contract contained some positive incentives, it sharply increased the total cost of in-hours GP services that had to be paid for by PCTs. Since many GPs opted out of out-of-hours services, these services had to be re-provided at a higher cost by PCTs. The new medical consultants' contract also resulted in large pay increases for medical consultants, with no obvious commensurate benefits for the NHS (Williams and Buchan 2006). The extra cost to the NHS was about £340 million per annum – £90 million more than was originally expected (Bosanquet *et al* 2005).

'Agenda for Change' is the name given to the pay reforms for non-clinical staff in the NHS. This scheme also resulted in a substantial increase in average pay costs for staff covered by the scheme. It required no change in working practices or increased flexibility as a quid pro quo for the pay increases. The extra cost of Agenda for Change when fully implemented was originally estimated to be £4 billion per annum (Royal College of General Practitioners 2002) but the actual cost is likely to exceed this.

A feature common to all these pay deals was that they were negotiated centrally, and PCTs and providers had no option but to adopt them. Pay increases were not linked explicitly to changes in working practices or higher productivity. All were expensive, and in each case the cost of the pay deal was underestimated, in aggregate by about £600 million (Jones 2006). PCTs and providers had to pay staff the full amount of the pay awards and absorb the funding shortfall, making it that much more difficult to achieve financial balance.

Declining hospital productivity

According to the Office of National Statistics, during the last decade there has been a steady decline in the productivity of NHS hospital providers, using the standard measure of outputs to inputs, amounting to 0.6–1.3 per cent per annum (BBC News online 2006b). There has probably also been a comparable decline in productivity in primary care as well, although data is too poor to be sure. It is true that some of the apparent decline in productivity is the result of failure of volume data to adequately reflect improvements in the quality of care. Nevertheless, declining productivity measured in this way means that it now costs more to provide the same volume of health care outputs than it did several years ago, even before taking account of cost inflation. Therefore a given NHS budget will purchase fewer services for patients.

Some of the causes of declining productivity were as follows.

- **Excessive focus on expanding inputs** The explicit focus was on increasing inputs, such as doctors, nurses, beds and so on, with little focus on getting more out of existing capacity – in other words, improving productivity. This was despite much evidence that, although there were certainly capacity bottlenecks in some places, there was no general shortage of all types of capacity everywhere. For example, for a long time there has been clear evidence that in London there is too much of certain types of hospital capacity and too many hospital beds (Tomlinson 1992).
- **New influences on working practices** The European Working Time Directive (EWTd), national service frameworks and new clinical protocols issued by several royal colleges all contributed in different ways to higher unit staffing costs and lower hospital productivity. Of course, in many cases these new influences on working practices are expected to improve the quality of clinical care – and in some cases they have already contributed to more flexible clinical working practices. Nevertheless, they increased unit costs and reduced productivity. The EWTd has had a particularly direct impact on costs in many clinical services, and additional EWTd costs will be incurred in 2009 when the current transitional provisions come to an end.
- **Stringent access targets** This factor also proved to be costly. In particular, the requirement to ensure that 98 per cent of A&E attenders had a maximum four-hour wait caused hospital trusts to incur high marginal costs to ensure full compliance with the target. The 98 per cent threshold imposed significantly higher marginal costs than would have been incurred if a slightly lower – say, 95 per cent – threshold had been set (based on experience of achieving the target at a major London teaching hospital). When the number of A&E attenders is rising and uncertain (which has been the case for many A&E departments in recent years), maintaining sufficient excess capacity to be 98 per cent certain that a patient can be moved from A&E within four hours is much more costly than being 95 per cent certain.
- **Low productivity and deferred productivity expenditures** There was a long list of new initiatives that had to be funded out of the additional NHS resources that either will never contribute to higher output of patient services, or will do so only in the future. One example is the creation of 300 PCTs in 2002, which added an additional approximately £1.5 billion per annum to the NHS budget just to fund the management overhead costs. Another is the National Programme for IT (NPFIT). Now expected to cost

£2 billion each year for 10 years, this will improve clinical services and the quality of patient care greatly and may improve productivity but not for some time. Irrespective of whether these expenditures represent value for money, they are likely to reduce productivity in the short and medium term, and to consume additional NHS resources, thereby reducing the volume of patient services that can be provided out of the available resources.

Inadequate policy levers

The fundamental cause of the loss of financial control of the NHS as a whole was the failure to put in place policy levers that were sufficiently effective to manage demand, to control cost inflation and to ensure that providers delivered sustained improvements in productivity.

On the demand side, practice-based commissioning (PbC) was only introduced as an afterthought. PCTs had very weak levers to manage demand and failed to contain activity growth to affordable levels. The rapid growth in non-elective activity was accommodated and funded, not managed. The extra cost of funding the growth of emergency attendances and admissions in 2005/6 is estimated to have consumed a high proportion of the entire available uncommitted funding increase after allowing for cost pressures.

On the supply side, the higher cost of pay settlements was significantly underestimated and underfunded. Since payment by results (PbR) tariffs covered only a small fraction of total activity, there were no adequate mechanisms in place to ensure that activity growth and productivity improvement went hand in hand. This was the case, despite the fact that Sir Derek Wanless had earlier stressed that achieving the desired health outcomes depended on achieving sustained productivity improvement (Wanless 2002, p 119). The supply side operated as a giant ‘cost-plus machine’, with PCTs funding rapidly growing demand at much higher unit costs.

At the system level, the failure to achieve NHS-wide financial balance should not have been a surprise given the weakness of the demand-management instruments, rapidly rising unit costs and inadequate retention of reserves. There was no reason to expect that aggregate expenditure would remain affordable. If the cost of providing total activity exceeded what was affordable, then NHS finances would not balance – and in 2004/5, and even more so in 2005/6, they did not.

The deficits incurred by more than 60 trusts in 2005/6 resulted in part from the differential impact of the staged introduction of the reforms – particularly PbR – and in part from the differential trust costs of meeting the waiting-time targets. There were bound to be ‘winners’ and ‘losers’ emerging during the transition from the ‘old’ to the ‘new’ NHS. Winners would find it easier to achieve financial balance, while the losers would be more likely to incur deficits and to have to adjust over time. The financial regime that required all trusts to maintain precise income and expenditure balance in every year was a mistake – it unnecessarily turned a period of necessary adjustment into a perceived financial crisis.

The apparent large size of reported trust deficits was exaggerated by the use of inappropriate resource accounting and budgeting (RAB) accounting, which distorted the apparent magnitude of the imbalances and made the task of restoring balance impossible

for trusts with large deficits. (For more on RAB accounting and why it is inappropriate, see Financing adjustment from the 'old' to the 'new' NHS, pp 42–3.)

Summing up

The financial crisis was not caused by 'too much reform, too quickly' as some would have it. This is clear because the main instruments of reform, such as PbR, PbC and patient choice only began to 'bite' in earnest in 2006/7. The real cause of the crisis is better described as 'too little reform, too late' – the result of rapid growth in funding before putting in place adequate levers to manage demand and to improve provider productivity.

In 2005/6, the extent of the financial problems confronting the NHS became increasingly clear. New policy initiatives, partly in response to the financial problems, were announced by the Department of Health. The question to which we now turn is whether this policy response was appropriate and whether the responses of commissioners and providers to the new policy initiatives can be expected to result in achievement of the government's wider health reform objectives as well as restoring the NHS to financial balance.

3

How is the NHS likely to respond to recent policy developments?

In December 2005, the Department of Health published its framework for NHS reform (Department of Health 2005a) in which it set out its intention to produce annual system rules for the NHS until the reforms are in place in 2008. The document acknowledged that further development of the reform instruments was needed – especially the commissioning framework and rules for system management and regulation – and further guidance on those topics was promised in 2006. In January 2006, the system rules to apply in 2006/7 were published (Department of Health 2006h), setting out the rules for that year, followed in February by the new payment by results (PbR) tariffs to apply in 2006/7 (Department of Health 2006d). In July 2006, the new commissioning framework was published (Department of Health 2006b). These recent policy developments are already having a major impact on primary care trusts (PCTs) and providers.

This section examines the following policy developments:

- 2006/7 system rules
- 2006/7 PbR tariffs
- the commissioning framework.

Having analysed the likely response of commissioners and providers to these policy developments, we will then consider whether the government's reform objectives are likely to be realised.

2006/7 system rules

The 2006/7 system rules can best be viewed as an urgent response to some of the causes of the financial crisis in 2005/6. Key provisions are summarised as follows.

■ **Strengthening levers to manage hospital demand – especially emergency admissions**

There will be full roll-out of practice-based commissioning (PbC) in 2006/7. This is intended to moderate the rate of GP referrals to hospitals. There are explicit targets to reduce emergency bed-days by 5 per cent and to limit the growth in aggregate elective activity from all providers (including the independent sector) to 3 per cent. There is a new instrument to more effectively manage demand for emergency admitted care. Emergency admissions above a specified baseline volume will be paid a lower marginal price equal to 50 per cent of the full tariff. PCTs have joint responsibility with providers for ensuring that elective demand is contained within affordable limits but PCTs are not permitted to cap elective volumes or pay a lower marginal price for GP referrals above planned levels.

- **Putting strong pressure on NHS hospital providers to improve productivity** In 2006/7, PbR tariffs will apply to about two-thirds of hospital activity. Average PbR tariffs increase by just 1.5 per cent in nominal terms despite expected cost inflation of 6.5 per

cent, so in real terms tariffs decline on average by 5 per cent. This reduction is justified as a 2.5 per cent efficiency factor and a further 2.5 per cent for ‘activity and cost adjustments’. The activity and cost adjustments will impact on providers differentially with early adopters of PbR (the first-wave foundation trusts) the hardest hit. There is also a rebalancing of the tariffs – elective tariffs increase on average by 5 per cent nominal while non-elective tariffs decline by 0.5 per cent nominal.

These measures are clearly intended to reverse the earlier pattern of declining hospital productivity and ‘force’ rapid substantial productivity improvements by year end. The large reduction in the real value of non-elective tariffs and the lower marginal price paid for volumes above baseline will reduce the incentive to increase emergency admissions above baseline and reduce the cost to PCTs of continued rapid growth in non-elective demand, if it occurs.

- **Imposing a tougher financial regime on PCTs and provider trusts** All trusts that were in financial balance at the end of 2005/6 are required to budget for a 1 per cent income–expenditure surplus in 2006/7. Trusts that were in deficit at the end of 2005/6 must plan to achieve recurrent balance and to repay by year end their legacy deficits, including resource accounting and budgeting (RAB) effects. The Department of Health ‘expects [that the system as a whole will] recover any overspend from 2005/6 and we are planning for a surplus’ (Department of Health 2006g, p 5). Where deficit trusts are unable to achieve these targets, they must, as a minimum, achieve recurrent balance, and their strategic health authority (SHA) may agree to give them a limited further short period to repay legacy deficits.
- **Top-slicing PCT allocations** All PCTs located in an SHA region with a deficit at the end of 2005/6 (regardless of whether the PCT itself had a deficit) are to have their funding allocations top-sliced, to create SHA reserves to ‘deal with legacy issues and to deliver service improvements’ (*ibid*, p 6) – in other words, to finance the 2005/6 deficit and any additional deficits incurred in 2006/7 and, if anything is left over, to finance service improvement.

Top-slicing reduces the net purchasing power of PCTs and, therefore, the affordable amount of services they can purchase for their patients. Top-slicing has a differential impact on trusts depending on whether they are located in a region with an overall deficit in 2005/6. In London – which was a deficit region in 2005/6 – top-slicing reduces PCT allocations by 3 per cent (about £300 million). Net of 3 per cent top-slicing and cost inflation, funding growth available to PCTs, after allowing for the benefit of the real reduction in PbR tariffs, is less than is required to fund the target 3 per cent elective activity growth. This means that if elective activity grows by the planned 3 per cent in their area, those PCTs are likely to incur a deficit.

- **Halving the purchaser parity adjustment (PPA)** The Department of Health had earlier put in place transitional arrangements to dampen the effects on PCTs and providers of the shift from locally negotiated prices to PbR tariffs. Some PCTs will have to pay higher prices to their providers under PbR than they were previously paying when locally negotiated prices applied. Those PCTs were to receive PPAs – compensation payments to offset the loss of purchasing power they would otherwise suffer. The system rules unexpectedly halved the PPA payments that certain PCTs had expected to receive in 2006/7, and gave notice that PPA will be phased out entirely by 2008/9.

Some PCTs located in deficit regions that are affected by both top-slicing and halving of the PPA will see sharp reductions in their net allocations, and the author's estimates suggest that they will be able to afford to purchase very little activity growth in 2006/7. Either they will have to find ways to suppress demand, denying patients treatment, or they will incur deficits.

- **Confirming that the second-wave independent diagnostic and treatment centres will go ahead** Payments to the independent sector reduce the amount of elective activity that PCTs can afford to purchase from NHS providers. Payments to first- and second-wave schemes are estimated to be £750 million in 2006/7 and £1 billion per annum thereafter, focused on a narrow range of elective procedures. The impact will be differential across NHS providers, with those located closest to the new centres being hardest hit. NHS providers whose commissioners are subject to top-slicing and halving of the PPA and are also located adjacent to independent sector treatment centres are likely to see an absolute reduction in demand for those elective services provided by the independent sector.

Implications of the 2006/7 system rules

The system rules are intended to restore NHS financial balance and deal with the deficits from earlier years. Top-slicing and withdrawal of PPA funding is being used to create SHA reserves to fund past and current deficits. This reduces PCT purchasing power, and therefore reduces the volume of services that PCTs can afford to purchase – for some PCTs, massively so. Top-sliced PCTs will have to manage demand very aggressively to keep expenditure within affordable limits. Although they now have stronger instruments to manage the cost of emergency admissions, they have only weak instruments to manage elective demand. If they do not succeed, they will incur deficits. If they do succeed, they will achieve financial balance at the expense of their providers, which will see much lower activity and income growth in 2006/7 than was expected and planned for last year.

The system rules are also intended to strengthen pressure on hospital trusts to improve productivity. The extension of tariffs to two-thirds of hospital activity, and the large reduction in the real value of tariffs, mean that 2006/7 is clearly destined to be a very tough year indeed for all hospital trusts. Hospital trusts – particularly those in deficit regions – will no longer be able to 'trade their way out' of deficits. They will be forced quickly to reduce unit costs without reducing the volume of activity that they provide. On average, hospital trusts must reduce unit costs by 4–5 per cent by year end, but deficit trusts will have to reduce them by a great deal more. If they succeed, the combination of higher productivity and lower activity growth will reduce the need for wards, beds and some staff, and stranded capacity is likely to increase. The quality of patient care may suffer at some trusts (Lloyd 2006). If they fail to reduce unit costs sufficiently, there will be another large aggregate provider trust deficit at year end.

The financial pressures will be highly differential across PCTs and hospital trusts. Those PCTs that are subject to top-slicing and halving of the PPA, and that are least well placed to manage hospital demand, will face the greatest financial pressures. Those hospital trusts that provide services in deficit regions, that are net losers under PbR, that are most adversely affected by independent sector provision and/or that started the year with a deficit will face the greatest financial pressures.

The NHS financial framework

For the first time, the 2006/7 system rules set out an explicit financial framework that recognises the affordability constraint on activity growth (Department of Health 2006h). PCT allocations before top-slicing will increase by £5.4 billion (9.2 per cent) in money of the day over the 2005/6 budget. Cost inflation is budgeted at £3.8 billion (6.5 per cent), leaving uncommitted resources before productivity gains of £1.6 billion. The Department of Health assumes that an additional £1.6 billion will be funded through efficiency improvements of 2.5 per cent across all expenditure categories – not just hospital activity (Department of Health 2006g, p 6). If this were achieved, it would give £3.2 billion (about 5.4 per cent growth) of uncommitted funding to pay for planned activity growth, to repay last year's deficit and to fund any additional deficits incurred in 2006/7, as well as paying for new service developments.

At first glance, it appears that the numbers do add up. However, on closer scrutiny they appear to depend on some very challenging assumptions. At the end of 2005/6, PCTs and provider trusts were overspending by about £1.2 billion in aggregate. If the gross trust deficit for the previous year has to be repaid in full (as the system rules state), then uncommitted extra funding is reduced from £3.2 billion to £2 billion – sufficient for PCTs to fund 3.4 per cent activity growth on average across all trusts.

In practice, the 2005/6 deficit will be repaid largely by regions that were in deficit at the end of 2005/6, so affordable activity growth for PCTs in those regions will be much lower than 3.4 per cent. Moreover, trusts in aggregate start 2006/7 in recurrent deficit, and so must reduce costs by another £1.2 billion just to get back into recurrent balance in 2006/7. Removing a recurrent gross overspend of £1.2 billion requires an average reduction in unit costs of not 2.5 per cent but 4.4 per cent. Trusts starting the year in deficit will have to reduce unit costs a great deal more than this average, some of them by as much as 10–15 per cent in one year. If trusts fail to achieve these target cost reductions in full, there will be a sizeable gross trust deficit again at 2006/7 year end.

Furthermore, the 2.5 per cent efficiency target is assumed to be achieved across all providers, whereas the levers to bring it about (PbR tariffs) apply only to about 35 per cent of total expenditure. If the 2.5 per cent efficiency saving were to be achieved only from tariff activity, then the funding released by efficiency improvements would be reduced by a further £1 billion, and uncommitted resources would fall from £2 billion to just £1 billion – well short of the funding increase required to pay for the expected growth of activity. The result would be either reductions in (non-tariff) services available for patients and/or higher trust deficits.

The year 2007/8 is the last year of committed 7 per cent real funding growth. If funding grows by 9 per cent nominal, then there will be an additional approximately £5.7 billion additional resources available in the hospital and community care budget (author's estimate). How much of that can be regarded as available to fund activity growth and service improvement will depend on:

- the rate of cost inflation and other unavoidable cost pressures
- whether there are outstanding legacy deficits at the end of 2006/7
- how much additional productivity improvement can be achieved in 2007/8.

For example, if cost inflation and other cost pressures could be reduced from 6.5 per cent in 2006/7 to 5 per cent (£3.25 billion) in 2007/8 and a further 2.5 per cent productivity

improvement could be achieved across all providers, then there would be an additional £3.5 billion (approximately) available before dealing with deficits – sufficient to fund on average 5.6 per cent activity growth.

However, if trusts end the year with a reduced but sizeable gross deficit – as currently seems much more likely given that at the end of the first quarter of 2006/7 the predicted gross trust deficit for the end of 2006/7 was £800 million (Department of Health 2006e) – then PCT allocations would be top-sliced again in 2007/8. For example, if the aggregate trust deficit was £800 million at 2006/7 year end, trusts would have to improve productivity next year by about 4 per cent across the board to restore balance, and affordable activity growth would be reduced to about 4 per cent after repaying the prior-year deficit. Affordable activity growth would be much lower for PCTs located in regions that started 2007/8 with a deficit. This means that whether activity growth required to achieve the 18-week wait target is affordable will depend on containing cost pressures, maintaining strong further productivity improvement, and the state of aggregate trust finances at the end of 2006/7.

From 2008/9, the NHS will see much slower funding growth. Statements by the Chancellor suggest that the most likely growth of NHS funding lies in the range of 3–4.5 per cent in real terms (HM Treasury 2006). With cost inflation and unavoidable cost pressures likely to consume more than half of that increase, additional funding will probably be sufficient to fund activity growth of only 1–1.5 per cent per annum unless further strong productivity improvement is sustained across the NHS.

Three conclusions seem clear. First, the 2.5 per cent efficiency factor in hospital tariffs looks likely to be here to stay beyond 2006/7. Second, if sufficient resources to support continuing health improvement are to be available, then ways will have to be found to apply similar pressure for productivity improvement on all other providers in primary care and mental health as well. Third, there is now an urgent need for commissioners to focus scarce health resources on achieving health care priorities. Commissioners need effective tools to ensure both that the money gets spent where it will do the most good, and that expenditure remains affordable.

Payment by results

In February 2006, the Department of Health published the new PbR tariffs for 2006/7. Fixed tariffs are to apply to about two-thirds of all hospital activity and the 5 per cent reduction in the average real price will put strong new pressures on hospital providers to reduce unit costs.

Intended effects

PbR is a vital piece of the reform armoury. Well-designed tariffs will force providers to reduce unit costs, thereby generating the additional resources needed to expand and improve services. Similar types of arrangements adopted, for example, in Australia, have induced major improvements in hospital lengths of stay and in day case rates, with outcomes that have been good for patients and cheaper to provide.

PbR tariffs would be cost reflective if they were set so that for each completed procedure the provider received an amount equal to the costs that they would have incurred if they

were operating efficiently. If PbR tariffs were cost reflective, then for the efficient provider, income will equal cost at the levels of health care resource group (HRG), service and trust. At the service level, the difference between income and costs would provide clinicians and managers with an accurate measure of their efficiency relative to other NHS providers.

Unintended effects

If tariffs are not cost reflective, then PbR can induce effects that are not only unintended but also undesirable. Unfortunately, there is an increasing amount of evidence that the 2006/7 tariffs are not cost reflective. The basis for this assertion is both analytical and based on an impact analysis of the 2006/7 tariffs.

The methodology for setting PbR tariffs in England is opaque and complex. There are at least 12 steps involved in deriving the tariffs from the reference cost data submitted by trusts. The most significant of these are as follows.

- Each trust reports its reference costs for the baseline year as total costs less income derived from activities other than patient services, including teaching and research and commercial activities. For some trusts, these activities reduce reported reference costs by as much as a third of total costs.
- Costs are allocated across procedures (HRGs) using the Department of Health cost-allocation rules. Activity is reported as finished consultant episodes (FCEs).
- Each trust's reference costs are deflated by its market forces factor (MFF) to convert reported reference costs to a common 'currency'.
- For services not funded by tariffs, PbR exclusions are deducted from deflated reference costs, and various other costs, such as NICE (National Institute for Health and Clinical Excellence) adjustments, are added in.
- The Department of Health converts FCEs to spells (the measure of activity used in computing the tariffs) using hospital episode statistics (HES) data to match FCEs to spells.
- Average costs per spell are computed for each HRG, using these deflated adjusted reference costs to derive uniform tariffs for the baseline year.
- Tariffs per HRG spell for the baseline year are inflated for two years of cost inflation and the efficiency factor, to derive uniform prospective tariffs pre-MFF inflation.
- Uniform prospective tariffs are adjusted by each trust's MFF to derive trust-specific prices for tariff services. (The MFF payments are not included in the tariffs but are separately funded.)

Concerns with the current approach

This approach raises a number of concerns. Three of the most significant are as follows:

- the assumption that the income received to fund non-patient service activities exactly equals the cost of providing those services
- the basis used for funding capital costs in the tariffs
- the basis on which the MFF is set is not cost reflective.

CONCERN 1: THE ASSUMPTION THAT THE INCOME RECEIVED TO FUND NON-PATIENT SERVICE ACTIVITIES EXACTLY EQUALS THE COST OF PROVIDING THOSE SERVICES

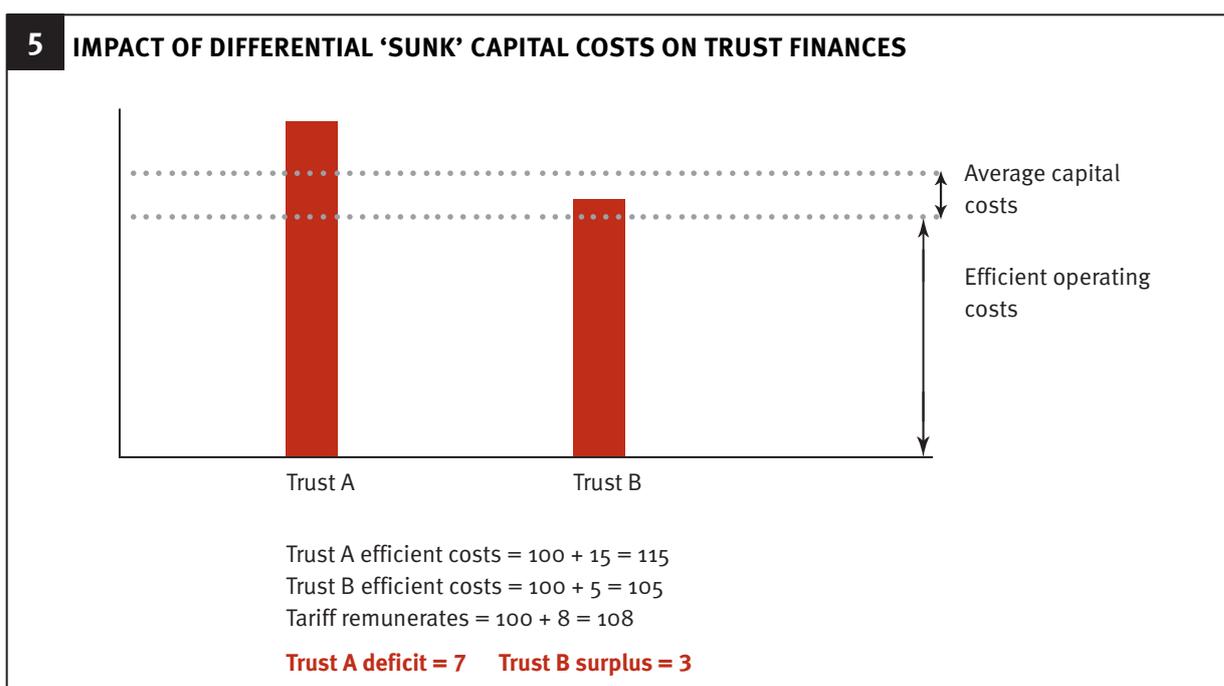
Since there are no reliable estimates of the costs of providing non-patient service activities such as teaching and research, the income received by each trust to fund those activities is deducted from reference costs rather than an estimate of the costs incurred

in providing them. The implicit assumption is that the income received to fund them exactly equals the cost of providing them. This is highly questionable and unlikely to be correct because that was not the basis on which funding was originally based. If it is not correct then a trust's reported reference costs for patient services will be too high or too low. Since some trusts have a lot of excluded activities and others very few, tariffs based on reported reference costs may not be cost reflective and some trusts will be overfunded at the expense of others.

CONCERN 2: THE BASIS USED FOR FUNDING CAPITAL COSTS IN THE TARIFFS

Trusts' total costs are made up of revenue costs (such as pay and drugs costs) and capital costs. For non-PFI (private finance initiative) hospitals, capital costs are the sum of depreciation of fixed assets and the dividend on public dividend capital (PDC). For PFI hospitals, they are the capital charge element of the annual payment to the PFI provider, which consists of depreciation and the cost of private finance. PbR tariffs are set to equal the national average of trusts' total costs. If the proportion of capital costs in total costs varies across trusts (for example, because one trust operates an old fully depreciated asset while another operates a new hospital building), then basing tariffs on the average of total costs will overfund hospital trusts with lower-than-average capital costs and underfund trusts with higher-than-average capital costs.

Figure 5 illustrates this point (see below). Trust A and Trust B both have recurrent costs equal to the national average, assumed to be 100. Trust A has capital costs relating to 'sunk' capital expenditure of 15, and Trust B has capital costs of 5. The national average of capital costs in this example is assumed to be 8. The tariff remunerates both trusts the same amount equal to 108. Trust A incurs a deficit of 7 and Trust B a surplus of 3, despite their controllable costs being exactly the same as one another and equal to the national average. The key difference between revenue costs and 'sunk' capital costs is that there is nothing a trust can do to reduce the latter. The income–expenditure charge will appear in the income–expenditure statement for the full life of the assets, so any unfunded element of this cost will appear as a continuing income–expenditure deficit.



This is not just a theoretical point. The Audit Commission has noted that the share of capital costs in total costs varies from 4 to 15 per cent across trusts in England, with an average of about 8 per cent (Audit Commission 2004). The current methodology for funding capital in tariffs systematically favours trusts with old capital stock, and underfunds those with new capital stock because new hospitals (however funded) are much more likely to have higher-than-average capital costs than older hospitals whose historic capital costs tend to be largely written down.

Most new hospitals are PFI schemes, so we should expect that hospitals with large PFI schemes will tend to be underfunded, and therefore more likely to incur deficits. This is exactly what Monitor observed when undertaking its recent foundation trust diagnostic work (Monitor 2006).

The PbR methodology particularly disadvantages early PFI schemes for a different but related reason: namely, that tariffs fund a 3.5 per cent return on public dividend capital, whereas early PFI schemes were approved and funded when the value-for-money test was that the cost of finance needed to be no greater than the then-public sector discount rate of 6 per cent. Funding in PbR tariffs sufficient to pay the current dividend on PDC will necessarily be less than the actual annual capital charge of all early PFI schemes. As a result, they will be underfunded and likely to incur deficits throughout the life of the PFI contract, even if they operate more efficiently than the average trust (Cambridge Economic Policy Associates 2006).

CONCERN 3: THE BASIS ON WHICH THE MFF IS SET IS NOT COST REFLECTIVE

The MFF is supposed to adjust the uniform national tariff to take account of non-controllable regional cost variations (Department of Health 2002), but it takes almost no account of the most obvious non-controllable cost variations, which are differential sunk capital costs. More than 90 per cent of the variation in MFF values is accounted for by variations across the country in an index of private sector pay rates. (The index weightings are 67.6 per cent staff, 4.6 per cent buildings and 0.6 per cent land, and 27.3 per cent is zero weighted so in practice only the private sector wages index is significant, accounting for 93 per cent of the variation in MFF values (Department of Health 2005b).)

Why this is regarded as an appropriate measure of non-controllable cost variations in the NHS is not clear. Pay rates for clinical and non-clinical permanent staff (which account for over 90 per cent of pay costs) in the NHS are based on national pay awards, with transparent regional weightings – for example, for inner- and outer-London living allowance. Although there may be some weak link between private sector pay rates and the cost of employing certain temporary staff, the related non-controllable regional cost variations are very small, and are directly observable. As the Department of Health guidance note correctly states, ‘Intuitively it should be possible to base the MFF directly on [variations in] the costs of employing NHS staff together with the additional (staff) costs such as agency staff’ (Department of Health 2005b, Annex A, para 3). However, this is not what is done. Instead, MFF values vary directly with differences in private sector pay rates, even though there is little correlation between these and the non-controllable regional cost variations actually incurred by trusts. The result is MFF values that are, by a long way, the highest in central London, and generally lowest in the north and west. (For a sample of MFF values for 2006/7, see Appendix 2, p 54.) (The Department of Health has argued that the approach can be justified because it is the same as is used throughout local government.

However, when a PbR regime and quasi-market are introduced, funding arrangements that were previously appropriate may need to be revised to take account of the new circumstances.)

Although, in theory, PbR sets a uniform national tariff for each procedure, in practice there is a different price for the same procedure for almost every trust, because the MFF values are different for almost every trust. Across the country, the difference between the highest and lowest price for the same procedure is about 50 per cent, which implies that it costs some trusts 50 per cent more to provide the same services as other trusts, solely because of non-controllable regional variations in pay costs. Moreover, variations in MFF values are large, even across small distances. For example, in London the MFF for St Mary's is 1.45, while for King's (less than 2 miles away) it is 1.29 and for Queen Elizabeth Woolwich (about 5 miles away) it is 1.20 (Department of Health 2006d). Is it credible that the costs for all procedures at St Mary's are 16 per cent higher than at King's because of 'regional variations in pay rates'?

The MFF values are of great importance because they are used to deflate trusts' reference costs to calculate national average costs and to re-inflate tariffs to derive trust-specific prices. Small differences in MFF values will have a major impact on trust income and financial balance. (For an example of this, see Appendix 2, p 53.) Small errors in setting the correct MFF value can cause trusts to have 'unintended' deficits or surpluses of as much as 5 per cent of turnover. It is also important to note that since the MFF is a scaling factor, if it is set too low for some trusts then it must have been set too high for others, resulting in undeserved transfers of income from 'losers' to 'winners' in ways that have nothing to do with their relative efficiency.

These flaws in the methodology used to compute PbR tariffs suggest that the computed tariff values are unlikely to be cost reflective. This proposition is supported by unpublished impact analysis of the 2006/7 tariffs undertaken by a consortium of NHS hospitals (McGuire personal communication 2006). Supporting evidence includes the following.

- A significant number of trusts have a reference cost index (RCI) greater than 100 but nevertheless are 'winners' under PbR. If tariffs were truly cost reflective and the RCI index was an accurate measure of relative costs, then all trusts with RCIs greater than 100 should be 'losers'.
- Differences in MFF-adjusted prices across trusts in a small area bear little relationship to estimated actual differences in non-controllable cost variations between those trusts.
- At the individual service level, tariff income bears no obvious or stable relationship to estimates of the expected costs of providing the service. Funding in the tariffs for some services is much more than any reasonable estimate of efficient costs, and for other services much less than is needed to fund efficient service provision. The adjustments to the tariffs for certain specialist services recently agreed by the Department of Health are an implicit acknowledgement that the tariffs for certain services are not cost reflective (Health Service Journal 2006c).
- Hospital trusts with early PFI schemes can demonstrate that a significant proportion of their deficit results from 'excess' sunk capital costs that are not funded in the tariffs (Cambridge Economic Policy Associates 2006).

DEFICITS DO NOT NECESSARILY IMPLY INEFFICIENT OPERATIONS

Each of the three factors noted here can have a large and differential impact on trusts. The net effect on some trusts may be small because the effects cancel one another out. However, the aggregate adverse effect on those trusts that lose out on all counts will be large. Equally, the benefit to trusts that gain on all counts will be large. Trusts that are overfunded will benefit at the expense of others, which must, therefore, be underfunded. Overfunded trusts that achieve balance will appear to be efficient while more efficient underfunded trusts may incur large deficits and appear to be inefficient. In reality, the allegedly efficient service may simply be benefiting from tariffs that are not cost reflective.

Learning from international experience

The methodology for setting PbR tariffs in England has been described as ‘superficially similar but essentially different’ (Walsh 2005) to the approach used in Victoria, Australia. Although the policy context in Victoria is very different to England, there are nevertheless relevant lessons to be learned from their experience with PbR. In Victoria, available funding is allocated strategically to specific health programmes, such as primary care, acute services and elderly care, to ensure that the money is spent in the way intended by the commissioner. Tariff prices are paid for baseline volumes of hospital care, but a lower marginal price is paid for volumes in excess of baseline. This enables the commissioner to manage the demand risk and limits the diversion of funding from primary into acute care.

Tariffs are based on average revenue costs only, and capital costs relating to ‘sunk’ capital costs are separately funded, to avoid the problems noted above that arise when tariffs are based on average revenue and capital costs. There is also separate funding to support clinical innovation. New or expanded services are tendered and the prices for those services are set by a competitive process. Table 2 summarises the key differences.

TABLE 2: DIFFERENCES BETWEEN PAYMENT BY RESULTS IN AUSTRALIA AND ENGLAND

Australia	England
Strategic funding allocations for programmes, such as acute services, elderly care, mental health, primary care and public health	No funding allocations by service or programme – money is spent wherever demand appears
Tariffs paid up to maximum volumes and lower marginal prices for extra activity	Tariffs paid for all GP-referred elective activity without limit
Tariffs based on recurrent costs only, with separate funding for capital costs	Tariffs based on average of recurrent and capital costs
Funding for clinical innovation outside of tariffs (transitioned into tariffs over time)	No separate funding for clinical innovation
Special tenders for new or expanded services – prices set by competitive process	No provision for pricing new and expanded services by competitive process or separate price from average cost tariffs

PbR in England

In summary, as currently designed, the PbR tariffs have two sorts of shortcomings. First, they provide perverse incentives that inhibit hospital trusts from supporting and participating in service reconfiguration, including a shift towards more care closer to home. Second, the unintended and undesirable effects caused by the non-cost reflective nature of the 2006/7 tariffs result in some trusts being underfunded at the expense of others that are overfunded. This causes some efficient trusts to incur deficits and to come under pressure to reduce costs below the efficient level with adverse potential consequences for their patients. Other less efficient trusts may achieve balance and avoid pressures to become more efficient simply because they are overfunded.

The commissioning framework

One of the key causes of the financial crisis was that PCTs had very weak levers to manage hospital demand. The 2006/7 system rules attributed great importance to full roll-out of PbC as the main mechanism for strengthening demand management and reducing inappropriate hospital admissions. The commissioning framework published in July 2006 (Department of Health 2006e) further clarified how commissioning is to work from now on.

In this section we consider the following:

- practice-based commissioning
- the new commissioning framework.

Practice-based commissioning

PbC is a key instrument in the armoury of reform. Essentially, it works by allocating GP practices an indicative budget based on baseline volumes of referrals, and allows them to retain a share of any savings resulting from fewer hospital referrals, or referrals to lower-cost providers. If well designed, PbC should improve the incentives acting on GP practices to help stem the flow of inappropriate hospital referrals and offer more appropriate care pathways for patients. The intention is that fewer patients will be added to hospital waiting lists, and more will be dealt with more quickly and cheaply by new services closer to home.

There can be little doubt that PbC is the right direction of travel for commissioning, putting clinical decision-makers at the heart of the commissioning process. However, in practice there are formidable challenges to be overcome before the potential of PbC can be realised. For PbC to be effective, two conditions must be met.

First, the incentives acting on GPs must be correctly calibrated. If the indicative budgets and benefit sharing are set with incentives that are too weak, then GP practices are unlikely to change their referral practices. However, if the incentives are too strong then GP practices will respond strongly but will keep most of the benefits, and PCTs will not be able to maintain or restore financial balance. Since there are no penalties for GP practices that fail to manage referrals down to an affordable level, the risk that demand exceeds what is affordable remains with PCTs. Getting the right balance is particularly problematic for PCTs that start the process with a deficit.

Second, if fewer hospital referrals are not to be achieved by denying patients care then there must be alternative referral options available to GPs, and they must be cheaper and more convenient for patients. At least for the next few years, there are unlikely to be many

such alternatives in many parts of the country, so in the short term PbC is unlikely to have a major impact on the flow of hospital referrals.

In due course, it is likely that new services closer to home will be created – some by GP practices, and some commissioned by PCTs. Whether the new services are cheaper for PCTs will depend on the cost of providing them and on whether the prices charged by providers of those services are cost reflective.

The 2006 commissioning framework

The commissioning framework describes commissioning as the process through which PCTs ‘secure the best value for patients and tax-payers’ (Department of Health 2006f, Annex, p 3). It states that every PCT is responsible for commissioning the full range of health services for its population. This implies that PCTs must make choices about where, and on what, their budget allocations will be spent. The description of the commissioning cycle explicitly includes ‘deciding priorities’ and ‘shaping the structure of supply’ (*Ibid*, p 6). Although PCTs have lead responsibility for commissioning, they are expected to devolve responsibility to GP practices via PbC.

A central theme of the commissioning framework is the importance of extending choice for elective services while recognising that in a cash-limited health system there will always be a need to set some boundaries around choice. Another potentially conflicting theme is the importance of PCTs managing demand to affordable levels to ensure their statutory duty to achieve financial balance is met.

The commissioning framework has a strong focus on the objectives of commissioning, but it is less explicit about how PCTs are to deliver them. PCTs now have stronger levers to manage the cost of emergency admissions because the split tariff reduces the cost of higher-than-planned admissions and reduces the incentive on providers to over-admit. But the instruments to manage elective demand are still weak.

For elective care, the commissioning framework retains the principle that hospital providers are paid full tariff for all activity provided. It sets out a range of possible demand-management measures that PCTs should consider adopting to avoid demand exceeding what is affordable. These include referral management centres, prior approval and utilisation reviews, which the framework states can be effective in managing demand ‘when used appropriately [and] where they carry clinical support’ (*Ibid*, p 23). However, experience in the United Kingdom and elsewhere suggests that scope to use these measures to manage demand in a clinically appropriate way may be limited (Health Service Journal 2006b; Donnelly 2006). In particular, referral management centres that seek to second guess the clinical judgements of GPs are likely to be unpopular with GPs and in conflict with the principles of PbC.

The commissioning framework recognises that the flow of referrals to hospitals may turn out to be unaffordable. Therefore, it states that ‘interim controls may... be needed’ and continues ‘Each year the Department of Health will publish an operating framework that... sets out the controls that can be appropriately applied’ (Department of Health 2006f, Annex, p 24).

From 2007/8, if hospital activity provided exceeds contracted volumes by more than an agreed margin, and if the reason for this is that a PCT has failed to manage GP referrals, then that PCT must pay the full tariff price for all activity provided. If activity exceeds the agreed threshold because of higher-than-planned consultant-to-consultant referrals, then PCTs may cap volumes and/or pay less than the full tariff for volumes in excess of the threshold.

These arrangements leave most of the risk that demand exceeds what is affordable with PCTs. Those PCTs that have seen their gross allocations reduced by top-slicing and halving of the PPA face a position where the level of GP referrals is likely to exceed what is affordable for them, yet they are obliged to pay the full tariff for all services provided – and they have a statutory duty to break even.

The commissioning framework also recognises that ‘a commissioning model based on choice and tariff alone may not always be sufficient to secure the development of new services’ (*Ibid*, p 3). It sets out a range of options that PCTs may use (if authorised by the SHA) to strengthen incentives for the provision of new services. The options include arrangements to pay new service providers a price supplement on the tariff, and/or a minimum income guarantee, and/or a share in the start-up capital costs.

Implications of the new commissioning framework

The rules set out in the commissioning framework still leave PCTs with weak instruments to influence the volume of activity that they must pay for, and the price that they must pay for it. This raises a number of important questions that the framework does not address, including the three questions set out below.

- **Question 1: How are PCTs to achieve financial balance when they have little influence over the volume of activity that they must pay for, and when they have to pay the full tariff for most of the elective activity provided?** In particular, how will PCTs that have been top-sliced and seen PPA halved be able to reduce patient demand well below the growth rate seen in recent years, and below the levels required to achieve the 18-week wait target, when they have weak levers to manage hospital referrals? The commissioning framework says that PCTs and providers should agree activity levels, but what happens if they cannot, or if they do agree them but the actual referrals exceed agreed levels?

NHS trusts and foundation trusts argue that there is a quasi-market, and that they should be free to use available capacity to provide more patient services if patients choose them. They also say they are obliged to make progress towards achieving the 18-week maximum wait, but this may involve an increase in elective activity that is unaffordable for PCTs and therefore inconsistent with PCTs’ statutory duty to break even.

- **Question 2: How are PCTs to shift funding from existing hospital providers to support provision of new services closer to home?** The framework recognises the need to fund providers of new services closer to home (and, in some cases, to offer them incentives), but does not address the fact that PCTs can only do this if they can be sure that spending on hospital care will be reduced. Since PCTs have no effective levers to bring about a reduction of revenue spending on hospital care, they are not in a position to contract with new providers to purchase new services because the new services will

only be affordable for the PCTs if hospital spending goes down, releasing funds to pay for them.

This problem is particularly acute for PCTs that are in deficit and/or under strong financial pressure because of factors such as top-slicing. PCTs cannot expect any help from hospital trusts to reduce hospital demand, because this will make it even more difficult for them to achieve financial balance. This results in a 'vicious circle' because if there were alternative cheaper services closer to home that GPs could refer to, then it would be much easier for GP practices and PCTs to manage hospital referrals. Since these do not yet exist, elective demand cannot be managed, so a high proportion of the PCT's budget is spent on hospital care, making it harder for them to afford to commission or provide alternative services closer to home. To address this dilemma, PCTs need stronger instruments to bring about a shift of funding from hospital care to new services closer to home.

- **Question 3: How are PCTs to deliver the commissioning priorities that they identify in their strategic plans for the medium term?** The description of the commissioning cycle in the framework clearly envisages a strategic commissioning role for PCTs, deciding priorities in collaboration with local stakeholders, and then shaping the structure of supply to ensure that the pattern of expenditure reflects those priorities. However, with the current weak commissioning regime, PCTs do not have effective instruments to achieve this. A PCT may want to spend more on disease prevention, cancer services or care closer to home – but if patients exercise choice for ever-more elective hospital care and all the PCT's budget is used up funding this care at full tariff, then there is little the PCT can do about it.

This sort of weak commissioning regime leaves the level and pattern of service provision largely 'to the market'. Demand is determined by short-term GP referral practices, and supply is determined by the response of existing providers to the incentives embedded in the tariffs (which are to maximise hospital admissions). In these circumstances, available PCT resources are used to fund greater use of existing capacity, leaving nothing for investment in new services. This already seems to be happening, with reports that PCTs are cutting their expenditure on prevention of illness and care closer to home in response to reduced net allocations and a rising bill for hospital services (Harding 2006; Martin 2006).

Over time, the demand for elective services and for drugs is always likely to outstrip available funding – particularly from 2008, when funding growth will slow sharply. In the absence of the ability of PCTs to strategically allocate resources in support of national and local health care priorities, most if not all of the available extra funding will be spent on more services from existing providers. The strategic shift of resources to support delivery of the PCTs' strategies is unlikely to be realised.

In summary, the weak commissioning regime will make it difficult for PCTs to deliver financial balance, difficult to effect a shift of funding from hospital care to support new services closer to home and difficult to deliver the commissioning priorities agreed locally.

Will the government's reform objectives be realised?

With the new system rules, the 2006/7 tariffs and the new commissioning framework now in place, can we expect that the government's reform objectives will be realised?

The outcomes of these initiatives will be heavily influenced by four factors:

- the weak commissioning regime, which has few effective levers for PCTs to manage demand and shift resources from lower priority services to higher priority services
- PbR tariffs that generate perverse incentives on providers to maximise admitted hospital care and that are not cost reflective and therefore overfund some trusts at the expense of others
- the absence, for the time being, of rules for system management and regulation of the 'new' NHS
- the difficult financial environment facing all provider trusts and many PCTs for the foreseeable future, which may well cause them to prioritise actions that will secure financial balance ahead of desirable service reconfiguration.

Slow progress reconfiguring services

As noted earlier, *Our Health, Our Care, Our Say* envisages a 'strategic shift' with 'more services.... closer to people's homes' (Department of Health 2006f). However, progress in bringing this about is likely to be slow for as long as PCTs have weak levers to make it happen, while hospital trusts will lose out financially if it does take place. Other government priorities include better care for patients with long-term conditions and more spending on prevention of illness. Progress in achieving the former will depend on changes to the tariffs to strengthen incentives on providers to minimise inappropriate hospital admissions, and to encourage them to provide more services in the community. Spending on public health and more care in the community is likely to continue to be crowded out by pressures on PCT budgets so long as all hospital care has to be paid for at the full tariff.

Despite the broad consensus around the idea that hospital providers should increasingly work collaboratively as part of networks of care, local agreement about reconfiguration of services to achieve this is also likely to be slow. This is because PCTs have weak levers to make it happen, some trusts are unlikely to be supportive because they will lose out in the short term and there are no adequate rules in place to manage and finance the restructuring process.

The 18-week wait

PCTs that have been top-sliced and have seen their PPA halved in 2006/7 will struggle to avoid year-end deficits. Achieving the 18-week wait by 2008 by expanding hospital activity may prove to be unaffordable for some PCTs – especially if trusts in a region start 2007/8 with a significant aggregate deficit and PCTs are again top-sliced to fund the deficits. Few PCTs in regions that are in deficit will be able to afford to commission or provide new services closer to home or sustain public health programmes.

'Undeserved' deficits

There will continue to be many trusts in deficit in 2007/8 and some of the deficits will be large. The distribution and magnitude of hospital trust deficits will not be a function of their relative efficiency but will depend on, among other things, the differential impact of the unintended and undesirable effects of the tariffs, the differential impact on PCT purchasing power of top-slicing and halving of the PPA, and whether the trust started the year with a deficit.

More 'stranded capacity'

Additional independent sector capacity will increase choice options for patients, but will also increase stranded NHS capacity for certain services. As the productivity of NHS providers increases and demand for hospital services flattens, stranded capacity will increase further. Very few new PFI hospitals will be required. The inflexibility of existing PFI contracts will constrain service reconfigurations and tend to bias solutions in favour of expanding activity at PFI hospital sites.

Restoration of NHS-wide financial balance

By the end of 2006/7, overall NHS financial balance will be restored if the aggregate recurrent and legacy deficit of all trusts is less than the value of reserves created by top-slicing and halving of the PPA. Current indications are that the reserves will not be sufficient and that overall balance may be achieved only by cutting public health, training and education spending. In future years, continuing top-slicing of PCT allocations is likely to be necessary if NHS financial balance is to be assured. This is because there will be a continuing need for sizeable reserves to fund unplanned growth in hospital demand paid at full tariff, should it occur, as well as to finance any remaining legacy and recurrent deficits and future restructuring costs.

Increasing health inequalities

Top-slicing and halving of the PPA risk turning equitable gross PCT allocations into inequitable net allocations. Health inequalities in some PCT areas may increase as a result, and primary care and mental health services may suffer at the expense of hospital providers.

Summing up

In summary, the new system rules, the 2006/7 tariffs and the new commissioning framework do address some of the causes of the financial crisis. They should help restore overall NHS financial balance, strengthen the ability of PCTs to manage the cost of higher-than-planned emergency admissions and strengthen incentives on NHS providers to reduce unit costs. However, the combination of weak commissioning instruments and perverse incentives embedded in the tariffs are likely to result in slow progress towards achieving other key government objectives, such as better care for patients with long-term conditions, more care closer to home, sustained improvements in the quality of hospital care and greater focus on the prevention of illness.

4

How can the reforms get back on track?

We have looked at the causes of the financial crisis in the NHS and have examined the recent policy initiatives introduced by the government in response to the crisis. The analysis suggests that, although overall system balance may be restored, progress towards achieving some of the government's key objectives is likely to be slow. In this final section we consider what further improvements to the design of the reform instruments would improve the prospects for more fully achieving the government's key health care objectives.

There are three key areas in which improvements to the reform instruments would generate renewed impetus towards full achievement of the government's stated objectives:

- **developing strategic commissioning** – a stronger commissioning regime that enables commissioners to shape the structure of supply more effectively in order to meet national and local health care priorities in the way envisaged in the commissioning framework
- **improving payment by results (PbR) tariffs** to make them more cost reflective, with improved incentives to support better care for patients with long-term conditions and more care closer to home
- **clarifying the system regulation and management rules** resulting in clear, workable rules, processes and institutions to regulate and performance manage the 'new' NHS.

This section addresses each of these points in turn.

Developing strategic commissioning

The commissioning framework document (Department of Health 2006b) says that primary care trusts (PCTs) will be expected to secure access to a range of high-quality services to meet local needs, including existing services and new services closer to home. The commissioning process is to include assessment of local needs, deciding priorities, developing a medium-term strategic direction for local services and shaping the structure of supply to ensure the best services for local people. The problem in practice is that the current weak commissioning regime does not give PCTs the instruments that they need to shape the structure of supply in the way envisaged. To do that effectively, they must be able to shift funding from low-priority services and re-allocate freed-up funds to support expansion and improvement of higher priority services.

What is strategic commissioning?

Here, the term 'strategic commissioning' is used to refer to a form of commissioning that gives PCTs more effective instruments to shape local services more actively in the way envisaged in the commissioning framework. Central to strategic commissioning are three-

to-five-year contracts between PCTs and providers to purchase specified ‘contract volumes’ of individual services or groups of services. The contract sets out contract volumes for each year of the contract. It also sets out rules about payments to be made if actual activity is above, equal to, or below the specified contract volumes. The payment arrangements are key. The current arrangement where the commissioner must pay the full tariff for all activity provided puts all the demand risk on PCTs, and makes it very difficult for PCTs to allocate resources according to local priorities. At the other extreme, an alternative arrangement would be that the commissioner only pays the tariff up to the specified contract volumes, and pays nothing for any excess, leaving the demand risk entirely with the provider and removing any incentive for the provider to expand activity beyond contract volumes. Such an arrangement would be incompatible with patient choice and would sharply increase the financial risks faced by providers, but it would enable PCTs to bring about a shift of spending from hospital care in favour of more care closer to home.

Strategic commissioning is the middle ground. PCTs would pay the full tariff price for contract volumes but a lower marginal price – say, 50 per cent of tariff – for any services provided in excess of contract volumes. This approach would replicate in elective care the approach adopted earlier this year for emergency admissions. It is compatible with supplier competition and patient choice because once the cap has been reached, all suppliers would continue to have the opportunity to provide additional services at the same marginal price. If the marginal price were set at 50 per cent of tariff, then there would continue to be an incentive for providers to offer additional services over and above the contract volumes.

This apparently small change in the terms of the PCT-provider contract has major implications for the way in which commissioning works. PCTs could bring about a shift in the allocation of resources over time, by changing gradually the contract volumes of different types of services. Reducing the contract volumes year on year that would be paid full tariff would signal that the commissioner was intending to reduce expenditure on that service and shift resources freed-up to fund other higher priority services. The provider could be expected to respond by seeking to reduce costs so that the service remained profitable – even if an increasing share of the services provided were paid less than full tariff – and if unable to do so to reduce the capacity of that service. This is an effective means of shifting resources out of hospitals and using the freed-up resources to create new services closer to home. It can also be used to bring about new networks of care that cut across existing institutional boundaries.

Strategic commissioning contracts can be used to contract new or expanded service capacity either for hospital services or for community services, or packages of care across the primary care/acute care boundary. They can be used to commission services from NHS providers and/or the independent and third sectors via a competitive tendering process. Once the new service capacity has been commissioned, GPs will have more referral options and patients will have greater choice.

Advantages of strategic commissioning

There are important advantages of adopting this approach to commissioning.

- Commissioners would have real ‘teeth’, and would be able readily to deliver their locally approved health care priorities by shifting resources to where they are needed.

- Hospital providers would be more likely to support the care-closer-to-home agenda because the incentives to expand hospital activity beyond contract volumes would be muted.
- Medium-term contracts of this type would considerably reduce provider uncertainty about the future affordable demand for hospital services, thereby facilitating local agreement on services reconfiguration.
- There would be less destabilisation of PCT- and NHS-provider finances associated with changes in the level and pattern of demand because activity in excess of contract volumes would be paid the marginal price (not average cost).
- Incentives on providers to improve efficiency would be greater because extra activity would only be profitable if marginal cost was lower than the marginal price (whereas previously it only had to be lower than the average cost).

Objections to strategic commissioning

Some may take the view that the level and pattern of services should be ‘left to the market’ rather than be set by commissioners in contracts. Unfortunately, as explained earlier, if one leaves things ‘to the market’ in a NHS where there is ‘weak’ commissioning, prices are fixed and there are perverse incentives acting on hospital providers to maximise admitted care, the result will not be the pattern of services sought by the government, PCTs or patients.

Critics could argue that paying the lower marginal price for volumes above the cap would reduce somewhat the incentive to expand supply. The response to this criticism is that affordable activity is strictly limited by PCTs’ fixed budgets, so it is not desirable to have very strong incentives to expand supply beyond the level that is affordable.

Strategic commissioning in the context of wider government objectives

The commissioning framework recognises that different commissioning approaches may be appropriate for different types of services. For example, it recognises that choice is less a driver of performance in non-elective care than in elective care but that it is of central importance for certain services, such as routine minor surgery, orthopaedics and certain diagnostic services. Similar important distinctions exist across other service areas as well. It is widely accepted that the best way to provide cancer or cardiac services is by operating networks of care with locally available outpatient services close to the patient, and centralised specialist inpatient services when required. In this model of care, patient choice should enable patients to choose between networks, not between hospitals. Similarly, treatment of long-term conditions is best managed across acute and primary care provider institutions with the aim of minimising inappropriate hospital admissions. Strategic commissioning provides a mechanism for developing effective networks of care (that may cut across the primary care/acute care boundaries) while preserving choice between networks.

In summary, strategic commissioning will enable PCTs to do more effectively the job given to them in the commissioning framework. It has greatest applicability to those elective services where provision is best managed across networks of care. It is supportive of practice-based commissioning (PbC) and patient choice and mitigates some of the perverse incentives embedded in PbR.

Improving payment by results tariffs

PbR is a cornerstone of the reform process – and rightly so. A well-designed PbR system can deliver many positive outcomes. However, the current process and mechanics for setting PbR tariffs are not fit for purpose – and left unchanged, PbR tariffs could do unintended harm. It would be a major mistake to conclude that PbR should be scrapped, but a complete overhaul of the process and mechanics is needed, and should be initiated immediately. (The Department of Health has recently announced a limited review of some aspects of the tariff setting that may refine the Department’s methodology but it is not expected to address most of the concerns set out earlier.)

The areas of work need to include:

- a new transparent and clinically based process for setting tariffs
- a review of tariff-setting principles and methodology
- incentives for quality improvement
- a decision about whether to use maximum prices rather than fixed prices.

Each of these is discussed below.

New transparent and clinically based process for setting tariffs

Following the saga of the issued, withdrawn, changed and then re-issued 2006/7 tariffs, few in the NHS doubt that the current process for setting PbR tariffs must be improved. Indeed, the Lawlor report (Department of Health 2006g), which reviewed the processes for setting PbR tariffs for 2006/7, confirmed the need for a major overhaul.

In other sectors, such as the utility industries, the United Kingdom is the home of best practice in price regulation. Best practice involves a transparent and extended process for price setting, including consultation with affected parties in a series of stages focusing sequentially on principles, methodology and data robustness, publication of interim findings, and draft determinations – all before final determinations are published. The process for setting PbR tariffs should draw on this best practice, and a similarly transparent and consultative process should be developed in health. Tariff design in health should include intensive and detailed engagement with clinicians. Given the complexity of the task, a process should be constructed that will lead to recommendations on tariffs for 2008/9. Any earlier target date is too soon, given the amount of work involved and the poor quality of data.

An independent body should be tasked with setting up and leading this process within clearly specified terms of reference set by the Department of Health. The independent body should make recommendations, but final decisions should remain with the Department of Health. (A government minister has been reported as agreeing with the view that an independent party should set or recommend the tariffs, although he stressed those were personal views (Harding 2006b).) The review should also address the transition issues involved in moving from the current non-cost reflective tariffs to the new tariffs.

Review of tariff-setting principles and methodology

The PbR review should address a range of issues relating to the principles and methodology for setting tariffs.

For episodic care it should address factors such as:

- the treatment of excluded costs from the reference cost basket
- the treatment of ‘sunk’ fixed costs and whether they should be excluded from the average cost calculation and be separately funded
- the methodology for allocating costs across health care resource groups (HRGs)
- the finished consultant episode (FCE)/spell conversion
- (in particular) the market forces factor (MFF).

A key concern should be to develop a revised methodology that is more likely to result in cost reflective tariffs.

The Department of Health has acknowledged that an activity-based tariff – and particularly one that encourages hospital admission – is not appropriate for treatment of long-term conditions where clinical practice should be focused on avoiding the need for hospitalisation wherever possible. There is a need to devise an alternative approach to payment for patients with long-term conditions that would reward the service provider for applying best clinical practice (with hospital admissions only when appropriate), and would retain incentives to improve productivity. Solutions will only be developed if there is a concerted effort involving clinicians as well as others.

The question of whether – and if so, how – PbR can be extended to mental health and primary care raises major questions to which there are currently no answers. In mental health, there are major reservations about applying PbR in its current form, with all its perverse incentives to maximise admitted care. In primary care, the lack of activity and cost data suggest that early adoption of PbR tariffs would be unwise. If in either of these areas a variant of PbR is not judged appropriate in the near term, there remains an urgent need to improve activity and cost information, and to devise payment arrangements that strengthen incentives on those providers to improve productivity in the same way as PbR has strengthened incentives on hospital providers.

Incentives for quality improvement

The commissioning framework has tabled for consultation outline proposals for strengthening incentives to maintain and improve the quality of care. Currently, the incentives to improve quality are weak – particularly in the many services in which patient choice is less directly a driver to improve performance. There is a risk that hospital providers faced with strong short-term financial pressures may prioritise financial balance ahead of efforts to sustain and improve the quality of patient care. This is the context in which consideration should be given to incorporating stronger quality improvement incentives into the PCT-provider contracts and PbR. In the opinion of the author, the options set out in the commissioning framework deserve serious consideration.

Maximum prices rather than fixed prices?

Because PbR tariffs are fixed, when stranded capacity emerges it sits idle, and hospitals with annual financial balance targets tend to close it down to save costs. In a market, the price would fall and the utilisation rate would rise as demand rose (because the price had fallen). If this were allowed in the NHS, then commissioners would be able to purchase more services for patients, and providers would be able to provide more services, thereby earning additional income and reducing their deficits relative to what they otherwise

would have been. This would mean less stranded capacity and fewer services closed – and everybody would be better off. Arguably the ‘new’ NHS can only work effectively when downward price flexibility is allowed (King’s Fund 2006a).

There is a strong case for seriously considering a gradual shift from fixed tariffs to fixed maximum tariffs with downward price flexibility – initially for activity in excess of contract volumes. If this were done, there would be stronger competition in supply, greater efforts by providers to further improve productivity, and more services available for patients within the fixed NHS budget. A useful start could be made by tendering the price for new or expanded services on this basis, that is, by setting a fixed maximum price but allowing bidders to offer a lower price if they chose to do so.

Clarifying the system-regulation and management rules

The Department of Health has recognised that there is a need to clarify the rules for regulating and managing the ‘new’ NHS. This clarification is now expected later in 2006. Here we consider the four key areas where clarification is needed:

- financing adjustment from the ‘old’ to the ‘new’ NHS
- regulation of providers
- performance management of commissioners and publicly funded providers
- financing service reconfigurations.

These four areas are discussed below.

Financing adjustment from the ‘old’ to the ‘new’ NHS

The transition from the ‘old’ to the ‘new’ NHS has caused some trusts to incur large deficits which, when Resource Accounting and Budgeting (RAB) effects (*see opposite*) are included, they will never be able to repay. At least part of their deficits are due to factors beyond their control, such as legacy costs, stranded costs and inappropriate MFF values (Palmer 2005b).

The 2006/7 system rules stated that all deficit trusts must seek to restore balance and repay past deficits in full (including the RAB effects) by the end of 2006/7. For some, it will be impossible to meet this target without harming the quality of their patients’ care. Recent statements by the Department of Health indicate that a few trusts will be granted a slightly longer time to restore recurrent balance. However, there is still little recognition that in a minority of cases the deficit is the result of large, unfunded costs that are beyond the control of trust management. Setting unrealistic financial targets for these trusts may only increase the risk that patient care will suffer and/or heighten the sense of failure if the targets are not met and thereby sustain the public perception that the NHS is in financial crisis.

This situation could be addressed by setting demanding but achievable medium-term deficit reduction targets for those trusts whose deficits are caused in part by factors beyond their control. The agreed deficit would taper over time, and would be financed with interest-bearing medium- or long-term loans either from the NHS Bank or from the strategic health authority (SHA) (funded out of the top-slice). This would extend the adjustment period over which deficit trusts were required to reduce controllable costs – first to restore recurrent balance, and then to repay the loans. Once a tapering ‘planned deficit’ had been

agreed with a trust, it would be permitted to report its financial performance by reference to whether or not the actual deficit exceeded the planned deficit.

This mechanism would not cost the NHS anything extra, since deficits have in any event to be financed out of SHA resources if the hospital is to remain open. This simple step would take the sense of financial crisis out of the adjustment process while retaining financial discipline on the trust to improve performance.

RAB accounting is not designed to deal with trusts that are subject to a prospective payment scheme. The way that RAB accounting works is that a trust deficit incurred in year 1 is treated as a liability to be repaid in year 2, and the amount of the deficit in year 1 is deducted from income in year 2, and in year 2 a surplus has to be generated sufficient to repay the year 1 liability.

As we can see in Appendix 3 (see pp 55–7), this means that, in effect, a trust that incurs a deficit of, say, £5 million in year 1 must reduce its costs by three times that amount – in other words, by £15 million – in year 2. This is because the trust starts year 2 with a recurrent deficit of £5 million, has its income in year 2 reduced by £5 million (giving a RAB-adjusted recurrent deficit of £10 million) and has to generate a £5 million surplus to repay the year 1 deficit.

Moreover, RAB accounting is cumulative – so if the trust in this example managed to reduce costs in year 2 by £5 million to restore recurrent financial balance before RAB effects, it would end year 2 with a cumulative RAB-adjusted deficit that is double what it was at the end of year 1. Hence in year 3 the trust would have to reduce costs by £30 million to clear the RAB deficit. We can see from this example that very quickly it can become impossible ever to clear the RAB-adjusted deficit.

RAB accounting does not give a true and fair view of the underlying financial performance of deficit trusts. Income–expenditure deficits should be treated as they are in other companies. The deficit creates a liability (in other words, a debt) in the trust’s balance sheet that is payable to the party that financed the deficit (in this case, the NHS Bank or the SHA). The charge in the income–expenditure statement relating to this liability should be the amount of interest accruing on the loan used to finance the deficit. (The recently published Audit Commission report on financial management in the NHS (Audit Commission 2006) has come to the same view about RAB accounting and has recommended that it should no longer be applied to hospital trusts. At the time of publication, the Department of Health was considering its response to the report.)

Regulation of providers

The outcome of the wider regulation of health care review is eagerly awaited. Among other things, it will need to set out rules to govern institutional restructuring transactions such as trust mergers, acquisitions, strategic alliances and franchising involving NHS trusts, foundation trusts and the independent sector. There will be instances in which locally agreed institutional solutions offer the best way forward to achieve the pattern of services that is best for patients and cheapest for the tax-payer. Such solutions cannot be developed by trusts until the rules that are to apply to these transactions are further clarified. Neither the principles foreshadowed in the system rules to test proposals for

provider mergers, nor the principles set out by Monitor in respect of mergers of foundation trusts, provide sufficient clarity to constitute a workable basis for formulating institutional restructuring proposals. (For more on regulation of health care, see Palmer 2005c.)

Performance management of commissioners and publicly funded providers

Whereas regulation will apply to all providers, performance management relates specifically to the role of owner's representative of publicly funded entities, such as PCTs, NHS trusts and foundation trusts. (Independent sector providers, of course, are performance managed by their shareholders.) The Department of Health has now clarified that SHAs will performance manage PCTs, holding them to account for the way they carry out their functions. The owner's representative for NHS trusts is currently their SHA, and for foundation trusts it is Monitor.

Currently, there is a lack of clarity around where regulation starts and where performance management of publicly funded providers ends. Who is responsible for assuring the quality of care provided by NHS providers – the PCTs, the Healthcare Commission, or both? Who is responsible for assuring the financial performance of NHS trusts and foundation trusts, and what do they do if a NHS trust or a foundation trust fails? What is failure of a PCT and how will it be dealt with? (For further exploration of these issues, see Palmer 2005b.)

Financing service reconfigurations

An important role of the owner's representative in future will be to recommend to the owner (the Department of Health) proposals to recapitalise trusts when needed, as part of desirable services restructuring. To effect service reconfiguration, it may be necessary to extend long-term loans, revalue public assets, defer the dividend on public dividend capital and/or refinance existing debt. Currently, there are no rules and processes to govern these matters, so service reconfiguration proposals are held up.

In summary, the 'new' NHS quasi-market can only work effectively when there are transparent and workable rules in place to govern its operation. Rules and processes are needed to address regulation of all commissioners and providers. The responsibility for ensuring that the rules are complied with and processes are followed needs to be clearly specified, and rules and processes to deal with the financial performance management of publicly owned providers are also needed.

Summing up

A stronger strategic commissioning regime that allows commissioners to allocate funding in a targeted way to deliver national and local health care priorities would much improve the prospects of achieving the government's health care objectives. Improvements in the incentives acting on hospital providers would increase the likelihood that they would enthusiastically support and participate in desirable services reconfiguration. Clear rules to govern the operation of the 'new' NHS are a pre-condition for a decentralised quasi-market to work properly. It is the author's view that adoption of the improvements proposed in this section will result in much fuller achievement of the government's stated health care objectives.

5

Conclusions

The central concern of this paper is whether the design of the health reform instruments is fit for purpose. Will the incentives generated by the reforms change the behaviour of commissioners and providers in ways likely to deliver the government's health care objectives?

This paper suggests that the 2005/6 financial crisis was the result of rapid funding growth before putting in place adequate levers to manage the growth of demand and to induce improvements in provider productivity.

We have seen that the new system rules and wider adoption of payment by results (PbR) tariffs address some of the causes of the financial crisis. They markedly strengthen pressures on hospital providers to reduce unit costs. They strengthen the ability of primary care trusts (PCTs) to manage the cost of higher-than-planned emergency admissions. And they create reserves out of which trust deficits will be financed. They should help restore NHS-wide financial balance but at the price of placing very strong financial pressures on PCTs and providers, with some much more affected than others. The risk is that the restoration of financial balance is achieved in some cases at the expense of the quality of patient care and more generally at the cost of slower progress towards achieving wider health care objectives.

We have also seen that PCTs still have only weak influence over the volume and cost of hospital care that they must fund. The combination of weak commissioning instruments and strong incentives on hospital trusts to maximise hospital admissions is likely to result in slow progress towards achieving key government objectives such as better care for patients with long-term conditions, more care closer to home, sustained improvements in the quality of hospital care and more focus on the prevention of illness.

A stronger strategic commissioning regime that allows PCTs to allocate funding in a targeted way to deliver national and local health care priorities would much enhance the prospects for achieving these objectives. The strength of such a regime would derive from giving PCTs much greater influence over the share of their budgets that are spent on elective hospital services.

A much improved process and methodology for setting PbR tariffs is urgently needed. In addition to improving the design of tariffs for episodic care, urgent attention must be given to designing tariffs better suited to funding the treatment of patients with long-term conditions, services provided across networks of care, services provided outside of hospital and services provided for mental health patients.

The 'new' NHS can only work effectively when there are rules and processes in place to regulate and manage health care providers. The development of workable, transparent rules and processes, and clarification of who is responsible for ensuring that they are followed, are urgent priorities.

The system rules aspire to create a system driven more by incentives to respond to patients than by top-down targets. The commissioning framework describes a world in which commissioners secure access to a wider range of services for patients. However, it is the author's opinion that, at present, the incentives work against achieving a number of the stated goals, and the commissioning framework is too weak to deliver them. The proposed improvements to the reform instruments offer the prospect of strengthening commissioning and improving incentives and thereby resulting in more rapid progress towards full realisation of these goals.

References

Appleby J (2006a). 'The NHS deficit problem has gone from bad to worse'. *Evening Standard*, 14 March. Available at: www.kingsfund.org.uk (accessed on 10 August 2006).

Appleby J (2006b). 'Where's the money going?' Briefing, 3 February. King's Fund website. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).

Audit Commission (2006). *Review of the NHS Financial Management and Accounting Regime, July 2006*. London: Audit Commission. Available at: www.audit-commission.gov.uk (accessed on 25 August 2006).

Audit Commission (2004). *Introducing Payment by Results: Getting the balance right for the NHS and taxpayers*. London: Audit Commission. Available at: www.audit-commission.gov.uk (accessed on 25 August 2006).

BBC News online (2006a). 'GPs "earn up to £250,000 a year"'. *BBC News online*, 18 April. Available at: <http://news.bbc.co.uk> (accessed on 25 August 2006).

BBC News online (2006b). 'NHS productivity rate falling'. *BBC News online*, 27 February. Available at: <http://news.bbc.co.uk> (accessed on 25 August 2006).

Bosanquet N, de Zoete H, Beuhler E (2005). *The NHS in 2010: Reform or bust*. London: Reform. Available at: www.reform.co.uk (accessed on 25 August 2006).

Cambridge Economic Policy Associates (2006). *Analysis of the Causes of the Queen Elizabeth Hospital Deficit*. London: CEPA. Due to be published September 2006.

Department of Health (2006a). *Chief Executive's Report to the NHS: June 2006*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006b). *Health Reform in England: Update and commissioning framework*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006c). *Implementing Payment by Results: Technical guidance 2005/06*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006d). 'Market Forces Factors 2006/07' in *National Tariff 2006/07: Payment by results*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006e). *NHS Financial Performance: Quarter 1, 2006-07*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006f). *Our Health, Our Care, Our Say: A new direction for community services*. London: The Stationery Office. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006g). *Report on the Tariff-setting Process for 2006/7*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2006h). *The NHS in England: The operating framework for 2006/7*. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2005a). 'Health Reform in England: Update and next steps'. Press release 2005/0445, 13 December. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2005b). *Payment by Results and the Market Forces Factor*. Supplementary technical paper to *Implementing Payment by Results: Technical guidance 2005/06*, published in 2004. London: Department of Health. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2004). *NHS Improvement Plan: Putting people at the heart of public services*. London: The Stationery Office. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2002). *Delivering the NHS Plan: Next steps on investment, next steps on reform*. London: The Stationery Office. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Department of Health (2000). *The NHS Plan: A plan for investment, a plan for reform*. London: The Stationery Office. Available at: www.dh.gov.uk (accessed on 25 August 2006).

Donnelly L (2006). 'Demand management "not a panacea"'. *Health Service Journal*, 29 June.

Farr M (2006). 'Data Briefing: Emergency admissions'. *Health Service Journal*, 2 March.

Harding M-L (2006a). 'Cuts loom as DoH struggles to fill £2bn gap in central funds'. *Health Service Journal*, 30 March.

Harding M-L (2006b). 'Independent body should set prices, says Warner'. *Health Service Journal*, 16 March.

Harrison A (2000). *Access to Elective Care: What should really be done about waiting times?* London: King's Fund.

Health Service Journal (2006a). "'Most challenging time ever" ahead'. *HSJ*, 22 June.

Health Service Journal (2006b). 'Referral systems "pissed off" US doctors'. *HSJ*, 22 June.

Health Service Journal (2006c). 'Specialist tariffs "topped up"'. *HSJ*, 20 July.

Healthcare Commission (2006). *Annual Report 2005/2006: Putting Patients First: A better experience of health and healthcare*. London: Healthcare Commission. Available at: www.healthcarecommission.org.uk (accessed on 25 August 2006).

Healthcare Commission (2005). *Annual Report 2004/2005: Our Progress: One year on*. London: Healthcare Commission. Available at: www.healthcarecommission.org.uk (accessed on 25 August 2006).

- HM Treasury (2006). *Releasing the Resources to Meet the Challenges Ahead: Value for money in the 2007 Comprehensive Spending Review*. London: HM Treasury. Available at: www.hm-treasury.gov.uk (accessed on 25 August 2006).
- Jones S (2006). 'Disbelief as Hewitt claims best year ever for NHS'. *Guardian Unlimited*, 24 April. Available at: www.guardian.co.uk (accessed on 25 August 2006).
- King's Fund (2006a). *Designing the 'new' NHS: Ideas to make a supplier market in health care work*. Report of an independent working group. London: King's Fund. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- King's Fund (2006b). 'King's Fund response to commissioning and patient involvement guidance'. Press release, 13 July. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- King's Fund (2006c). 'True Scale of NHS deficits masked, says King's Fund'. Press release, 7 June. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- King's Fund (2005). *An Independent Audit of the NHS Under Labour (1997–2005)*. London: King's Fund. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- Le Grand J (2003). 'Models of public service provision: command and control, networks or quasi-markets' in *Public Services Productivity*, pp 15–18. London: HM Treasury.
- Le Grand J (1993). *Quasi-markets and Social Policy*. London: Macmillan.
- Lewis R, Appleby J (2006). 'Can the English NHS meet the 18-week waiting list target?' *Journal of the Royal Society of Medicine*, vol 99, pp 10–13.
- Lloyd I (2006). 'Patient care is "at risk"'. *Health Service Journal*, 30 March.
- Martin D (2006). 'Wanless: public health reform has fallen victim to pay rises'. *Health Service Journal*, 6 April.
- McIntosh K (2006). 'Funding row could see 1,800 patients rejected by foundation'. *Health Service Journal*, 29 June.
- Monitor (2006). *NHS Foundation Trusts: Preliminary results for the year ended 31 March 2006*. London: Monitor. Available at: www.monitor-nhsft.gov.uk (accessed on 25 August 2006).
- National Leadership Network (2006). *Strengthening Local Services: The future of the acute hospital*. Report of the National Leadership Network Local Hospitals Project. London: National Leadership Network. Available at: www.nationalleadershipnetwork.org (accessed on 25 August 2006).
- Palmer K (2005a). *NHS Reform: Will it deliver the desired health outcomes?* King's Fund website. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- Palmer K (2005b). *How Should We Deal with Hospital Failure? Facing the challenges of the new NHS market*. London: King's Fund. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).
- Palmer K (2005c). 'Economic Regulation of the Health System in England'. Discussion note submitted to Currie review of wider regulation of health care. London: Cambridge Economic Policy Associates. Available at: www.cepa.co.uk (accessed on 25 August 2006).

Palmer K (2002). *Foundation Trusts and the New NHS Architecture*. Cambridge Economic Policy Associates website. Available at: www.cepa.co.uk (accessed on 25 August 2006).

Royal College of General Practitioners (2002). 'Agreement on NHS pay reform'. *The Week in Parliament*, 25 November–1 December 2002. RCGP website. Available at: www.rcgp.org.uk (accessed on 18 September 2006).

Tomlinson B (1992). *Report of the Enquiry into London's Health Service, Medical Education and Research*. London: HMSO.

Walsh M (2005). 'Output-based funding in England and Australia: superficially similar but essentially different'. Presentation to the King's Fund Market Futures Breakfast Discussion, December 2005.

Wanless D (2002). *Securing Our Future Health: Taking a long-term view*. Final report. London: HM Treasury. Available at: www.hm-treasury.gov.uk (accessed on 25 August 2006).

Williams S, Buchan J (2006). *Assessing the New NHS Consultant Contract: A something for something deal?* London: King's Fund. Available at: www.kingsfund.org.uk (accessed on 25 August 2006).

Appendix 1: Financial balance in the NHS

This appendix sets out more fully the activity, cost and funding relationships in the new NHS. It shows the determinants of affordability of activity and, in particular, clarifies why improvements in provider productivity are an important determinant of the volume of services that can be afforded, given any specified growth in health funding.

Each year, primary care trusts (PCTs) spend their allocations on purchasing services from primary care providers (in other words, GPs), mental health trusts and hospital trusts. They also provide certain community care services themselves. If PCTs are to be individually and in aggregate in balance in any year, then:

$$B = Q_p.P_p + Q_h.Th$$

where:

- **B** is the available budget,
- **Q_p** is the quantity of services provided by primary care providers
- **P_p** is the implicit price of primary care activity – meaning the amount of money spent divided by the quantity of services provided (which equals the average cost of providing the services)
- **Q_h** is the quantity of services provided by hospital providers
- **Th** is the average tariff for hospital services (which equals the average cost of providing those services).

The formula is a simplification, which ignores mental health and the fact that some hospital activity is not yet included within payment by results (PbR). However, these simplifications do not affect the analysis or the conclusions.

What the formula says is simply that to achieve financial balance, the budget that is spent by PCTs on services must equal the sum of their purchases, which consist of purchases from primary care (the quantity of primary care services times their average cost per unit) and purchases from hospitals (the quantity of hospital services times the average tariff for those services).

Important conclusions follow from this formula, which are discussed earlier (see pp 6–10) and explored more fully in Palmer 2005a.

The importance of productivity improvement

Another way of looking at affordability is to ask how much extra activity PCTs can afford to purchase year on year, given the announced increases in their funding allocations. Starting

from balance, the extra expenditure on all services must equal individually and in aggregate their extra allocations so:

$$|B| = |Qp|.Pp (1 + cp - pp) + |Qh|. Th (1 + ch - ph)$$

where:

- **|B|** is the percentage increase in the PCT allocation over the previous period
- **|Qp|** and **|Qh|** are the affordable percentage increases year on year in primary care and hospital care, respectively
- **Pp** and **Th** are the prices (= average costs) of primary care and hospital care, respectively, in the previous period
- **cp** and **ch** are the annual rate of cost inflation in primary care and hospital care, respectively
- **pp** and **ph** are the annual productivity improvements in primary care and hospital care, respectively.

Restating the same relationship in words starting from balance, and assuming no change in the share of hospital activity in total activity, we can say that:

$$\text{growth in affordable activity} = \text{growth in funding} - \text{cost inflation} + \text{productivity gains}$$

For example, if funding were to increase 6 per cent, cost inflation was 4 per cent and productivity improved 2 per cent, then:

$$\text{growth in affordable activity} = (6 - 4 + 2) = 4\%$$

However, if in this example cost inflation was 6 per cent and there was no productivity improvement then, despite the fact that PCT allocations increased 6 per cent, they would not be able to afford to purchase any more services for patients than in the previous year.

Appendix 2: The market forces factor

The market forces factor (MFF) value is used to deflate reference costs to calculate national average costs and to re-inflate tariffs to derive trust specific prices. Small differences in the MFF values will have a major impact on trust finances. The box below illustrates this point. In Case 1, the MFF is correctly set, and remunerates efficient trusts at levels that equal their costs. In Case 2, the MFF is wrongly set, and Trust A incurs a deficit and Trust B a surplus, even though they are of equal efficiency although subject to different non-controllable regional variations in costs.

IMPACT OF INCORRECT SETTING OF MFF

Case 1: MFF correctly set

Trust A actual costs = 120

MFF = 133

Trust B actual costs = 90

MFF = 100

Trust A deflated costs = $120/1.33 = 90$

Trust B deflated costs = $90/1.00 = 90$

Average national cost = $(90 + 90)/2 = 90$

Funding in MFF-adjusted tariff

Trust A = $90 (1.33) = 120$

Deficit = $120 - 120 = 0$

Trust B = $90 (1.0) = 90$

Deficit = $90 - 90 = 0$

Total funding all trusts = total costs = 210

Case 2: MFF not correctly set

Trust A actual costs = 120

MFF = 125

Trust B actual costs = 90

MFF = 100

Trust A deflated costs = $120/1.25 = 96$

Trust B deflated costs = $90/1.0 = 90$

Average national cost = $(96 + 90)/2 = 93$

Funding in MFF-adjusted tariff

Trust A = $93 (1.25) = 116$

Deficit = $116 - 120 = -4$

Trust B = $93 (1.0) = 93$

Surplus = $93 - 90 = +3$

Total funding all trusts = $116 + 93 = 209$

Source: Author calculations

The box below shows sample MFF values for 2006/7, with the lowest values (1.0–1.1) on the left of the figure and the highest values on the right.

MFF VALUES FOR A SAMPLE OF NHS TRUSTS AND FOUNDATION TRUSTS

West Cornwall 1.00	Manchester 1.10	Hillingdon 1.219	Hammersmith 1.302	Moorfields 1.40
Devon 1.023	Birmingham 1.119	Lewisham 1.23	Homerton 1.311	UCLH 1.42
Barnsley 1.047	Cambridge 1.139	Slough 1.235	Kensington 1.374	St Mary's 1.446
Norwich 1.05	Dartford 1.14	NW London 1.243		
Derby 1.059	Bexley 1.148	Brent 1.256		
Bolton 1.06	Oxford 1.17			
Bradford 1.06	Newbury 1.176			
Newcastle 1.06	Surrey/Sussex 1.183			
Sheffield 1.06				
South Tees 1.078				
Leeds 1.094				

Source: Department of Health 2006d

Note: 2005/6 MFF rebased to 1.00 in Cornwall

Appendix 3: How RAB accounting works

Resource accounting and budgeting ('RAB accounting') was introduced across the public sector to impose stronger controls on public spending. RAB accounting rules are complex. This appendix focuses on their impact on the income and expenditure accounts of hospital trusts.

Under RAB accounting, any overspend in a year is deducted from income in the following year, and the prior year's deficit must be eliminated by creation of an income–expenditure surplus in the following year after deducting income equal to the prior year's income–expenditure deficit. The income–expenditure surplus generated in the following years is then added to the following year's income.

The mechanics of RAB accounting are illustrated in a simple example in Table 3. In year 1, the trust is assumed to incur an income–expenditure deficit of 1. This creates a negative income–expenditure reserve on the balance sheet in that year. In year 2, income is reduced by the amount of the negative reserve, and the trust is required to achieve a surplus – in this case, of +1, whereupon the income–expenditure reserve is eliminated. In year 3, the trust is 'rewarded' with a one-off increase in income equivalent to the prior year's surplus, if in that year it has returned to recurrent balance.

TABLE 3: ILLUSTRATION OF RAB ACCOUNTING (1)

Year	1	2	3	4
Recurrent income	100	100	100	100
RAB income effect	–	–1	1	–
Recurrent expenditure	101	98	100	100
Recurrent income–expenditure balance	–1	1	1	0
RAB income–expenditure reserves	–1	0	0	0

Source: Based on figures from Cambridge Economic Policy Associates 2006

Note that in year 2, recurrent expenditure has to be reduced by three times the deficit in the previous year (from 101 to 98) to achieve this. The 'reward' in year 3 for removing the deficit incurred in year 1 is actually 'giving back' half of the previous year's recurrent surplus.

Table 4 illustrates RAB accounting for a case where in year 1 the trust has a large deficit (5 per cent of recurrent income) and succeeds in all following years to achieve recurrent income–expenditure balance before RAB effects. In year 2, despite having reduced expenditure by 5 per cent of income to achieve the recurrent income–expenditure balance before RAB effects – the RAB deficit doubles to –10 (10 per cent of income). In years 3 and 4, the deficit again doubles. It quickly becomes impossible for the trust to restore balance after RAB effects without sharply reducing activity, which under payment by results (PbR) reduces recurrent income and would increase the deficit further.

TABLE 4: ILLUSTRATION OF RAB ACCOUNTING (2)

Year	1	2	3	4
Recurrent income	100	100	100	100
RAB income effect	–	–5	–10	–20
Recurrent expenditure	105	100	100	100
Recurrent income–expenditure balance	–5	–5	–10	–20
RAB income–expenditure reserves	–5	–10	–20	–40

Source: Based on figures from Cambridge Economic Policy Associates 2006

RAB accounting was not designed to deal with entities that operate in a quasi-market where they are not able to fully predict and manage income, and where many of their costs are not controllable. Table 5 illustrates how income–expenditure deficits would be dealt with if they were treated in a similar way to temporary deficits in the private sector. In Table 5, it is assumed, as in the previous example, that the underlying recurrent income–expenditure position is a deficit of 5 in year 1, followed by income–expenditure balance (before deficit financing effects) in each following year. Here, it is assumed that the year 1 deficit is financed by borrowing repaid over the following three years.

The income effect in the income–expenditure account is an increase in interest payable – in this case, a 5 per cent interest rate applied to the outstanding borrowing at the start of the period. It is assumed that the new rule is that expenditure must be reduced to the level that would generate a surplus sufficient to finance the interest and repay the loan over three years. Recurrent expenditure needs to be reduced from 100 to 98.75 in year 2, 97.8 in year 3 and 97.9 in year 4. Thereafter, the trust would have reduced its costs to the extent that it would incur a surplus (in this case, of about +2.1) once the borrowing had been repaid.

TABLE 5: FINANCING OF INCOME–EXPENDITURE DEFICITS

Year	1	2	3	4
Recurrent income	100	100	100	100
Recurrent expenditure (before financing effects)	105	100	100	100
Recurrent income–expenditure balance	–5	0	0	0
Borrowing to finance deficit	+5	–	–	–
Repayment of principal	–	1	2	2
Income–expenditure impact of financing (interest on loan)	–	–0.25	–0.2	–0.1
Recurrent expenditure (after financing effects)	–	98.75	97.8	97.9

Source: Based on figures from Cambridge Economic Policy Associates 2006

King's Fund

11-13 CAVENDISH SQUARE
LONDON W1G 0AN

Telephone 020 7307 2400
www.kingsfund.org.uk

THE KING'S FUND IS AN INDEPENDENT CHARITABLE FOUNDATION WORKING FOR BETTER HEALTH, ESPECIALLY IN LONDON. WE CARRY OUT RESEARCH, POLICY ANALYSIS AND DEVELOPMENT ACTIVITIES, WORKING ON OUR OWN, IN PARTNERSHIPS, AND THROUGH FUNDING. WE ARE A MAJOR RESOURCE TO PEOPLE WORKING IN HEALTH AND SOCIAL CARE, OFFERING LEADERSHIP DEVELOPMENT PROGRAMMES; CONFERENCES, SEMINARS AND WORKSHOPS; PUBLICATIONS; INFORMATION AND LIBRARY SERVICES; AND CONFERENCE AND MEETING FACILITIES.

In recent years, the NHS has seen the most sustained period of funding growth ever. But despite the increased funding, the NHS is in deficit. In 2005/6, NHS trusts overspent by more than £1.2 billion and the NHS as a whole overspent by more than £500 million. This discussion paper looks at the causes of the NHS deficit in 2005/6. It then considers three recent policy developments – the 2006/7 system rules, the new payment by results tariffs and the commissioning framework – and asks what the impact of these policy developments could be and how they might be improved.

ISBN 1-85717-552-2



9 781857 175523