NHS payment reform: lessons from the past and directions for the future

Policy response

Anita Charlesworth, Loraine Hawkins and Louise Marshall

February 2014
The NHS faces an unprecedented challenge to bridge the gap created by demand and cost pressures outpacing funding growth. This has led to an increasing focus on the role of the NHS payment system in improving productivity and system efficiency.

The payment system has been looked to as a tool to help meet this challenge, and it certainly matters. There is some evidence that the system by which NHS commissioners allocate money to hospitals and other providers can have an impact on the quality and efficiency of health services, both in the NHS and other countries – although the evidence suggests that these effects are modest and difficult to sustain. The payment system is unlikely to be the right lever to use to drive major service change. Following the Health and Social Care Act, Monitor and NHS England have joint statutory responsibility for the design of the payment system and for setting prices. They have begun to examine the options for the future, and more radical reform of the payment system.

We welcome the focus on long-term reform, but our review of the evidence also counsels the need for both caution and realism – payment reform is not a quick fix. The lead times for new approaches are long. Neither is it likely to be a panacea. The effects of previous reforms on efficiency and quality have been modest. Good intentions can also have unintended consequences – paying for quality may seem axiomatically good, but the evidence from pay-for-performance schemes suggests that designing schemes which effectively link payment to quality or outcomes goals is very difficult.

This policy response, which presents recommendations for the longer-term development of the NHS payment system, is published alongside a broader review of the evidence, our research report *The NHS payment system: evolving policy and emerging evidence*. 
Key Points:

To ensure that the long-term reforms to the payment system support the NHS to deliver efficient, high-quality care, Monitor and NHS England need to develop an approach that has the following core features:

**Comprehensive** – reforms to the payment system have traditionally focused on a limited number of individual services, care settings or sectors. If the NHS is to improve the efficiency of the health system as a whole, it needs payment systems that cover the continuum of care and which create incentives for providing the ‘right care in the right setting’.

**Focused** – the payment system should be used to incentivise aspects of care for which there is evidence that payment systems are an effective lever (these being provider efficiency, transparency and accountability).

**Aligned with wider system changes** – it is also important to recognise the limitations of the payment system as a lever for change. To be effective, the payment system must be part of a wider package of aligned incentives, regulation, data and information improvements, plus investment.

**Transparent and evidence-based** – incentive schemes can only work if the organisations and clinicians whose behaviour they are trying to change understand what is required. Too often, the incentives are blurred or inconsistent.

**Predictable and credible** – delivering productivity improvements and service innovation takes time, so health care organisations need to be able to plan with greater clarity about the financial environment.

Drawing on the research evidence, we suggest that Monitor and NHS England should prioritise longer-term reform of the NHS payment system which:

- fundamentally reforms the payment system for emergency care, shifting away from Payment by Results towards global budgets for capacity, informed by standardised benchmarks
- develops nationally priced currencies for a wider range of non-emergency services, including community health services, to make it easier for clinical commissioning groups to commission bundles or packages of care for pathways or patient groups spanning hospital- and community-based services
- considers alternatives to funding the cost of capital that is currently part of the Payment by Results system.

These reforms will take time to design and time to implement. Their impact would probably not be felt for several years, even if there was significant commitment to reform and redesign of the payment system. In the shorter-term, our analysis suggests four areas where more rapid progress could be made:

- refocusing the pay-for-performance elements of the payment system towards improving the integration and coordination of care
• proactively engaging with the design and evaluation plans for payment pilots that support integrated care and improved outcomes (such as capitation for frail older people in Cambridgeshire and Peterborough, or Bedfordshire’s year-of-care contract for musculoskeletal services), to maximise the potential to generate design and implementation lessons for the wider NHS

• improving the efficiency incentives in the payment system through significant extension of best practice tariffs to more planned care

• improving the evidence base to determine the efficiency factor to underpin pricing decisions.

Alongside these reforms, payment systems can only function effectively if they are based on accurate cost information, so there needs to be sustained focus on improving costing information across the NHS. Incentives can only operate if they are clear, consistent and transparent. Stabilising the system so that the NHS has some certainty for planning and investment is also important.
Introduction

The NHS in England spent more than £110 billion in 2012/13 (Department of Health, 2013a). This spending covers a huge range of health care services from preventive to palliative, and routine to emergency. Services are provided by a large number of diverse providers in varied settings across the country, including general practice, hospitals and community nursing services. The balance of spending across these services must be optimised if the appropriate services are to be available to population groups in an equitable manner. In attempting to ensure that the limited NHS budget goes to the right places, a fixed annual budget is allocated to the commissioners of health care (clinical commissioning groups (CCGs) in local areas and NHS England at regional and national levels), taking into account the relative need for care of each population group. NHS commissioners determine the volume and mix of services to purchase for their population group, agree contracts for those services and pay the providers (which may be NHS, voluntary or independent sector organisations). How and how much commissioners pay providers for NHS-funded services is determined by the payment system.

The last ten years have seen rapid change and innovation in the English NHS payment system (Marshall and others, 2014). Much of this has been contentious and there is considerable debate about how the NHS should pay providers for care in future. The Health and Social Care Act 2012 implemented a further series of reforms and, for the first time, the NHS payment system is governed by a statutory regulatory framework. Following the Act, Monitor (as economic regulator) and NHS England have responsibility for the NHS hospital and community health services (HCHS) payment system, and 2014/15 is the first year in which Monitor and NHS England will set prices and other payment rules. The new statutory framework under which Monitor and NHS England operate covers:

- specifying the units or bundles of health care services for which a payment is made – known as currency specification
- setting national prices for some health care services and rules for pricing services without a national price – known as price-setting
- approving agreements between providers and commissioners to modify the national price where services are uneconomic at the national price – known as tariff modifications
- setting rules for local variations to the payment system when providers and commissioners agree different payments for new services or new approaches to the bundling of care – known as tariff variations.

This regulated payment system applies to around £70 billion of NHS funding, of which £30 billion is purchased using nationally set prices (referred to as the Payment by Results system) and £40 billion through locally negotiated contracts. Some contracts – for
example, mental health – have nationally mandated currencies but prices are agreed locally. For other services – community nursing, for example – both currencies and prices are set locally (Monitor and NHS England, 2013b).

These payment arrangements do not cover the payments to GPs, pharmacists or dentists for primary care services. However, payment systems for all these services are determined nationally – usually with a statutory underpinning – and are more rigorously and consistently applied than those in the so-called ‘regulated system’ for hospital and community health services. Monitor has no mandate over these payments; NHS England alone is responsible for negotiating payments with the professional associations.

Monitor and NHS England recognise that the payment system needs review (Monitor and NHS England, 2013a) and have begun to engage the NHS, researchers and commentators in discussions on the future direction of reforms to the payment system. To contribute to this debate, this policy response provides an overview of the recent evidence on elements of the payment system, which is described more fully in our accompanying evidence review, *The NHS payment system: evolving policy and emerging evidence* (Marshall and others, 2014). Drawing on this evidence, we offer a series of recommendations for Monitor and NHS England for longer-term reform of the payment system to better support the needs of the NHS as it faces increasing challenges from austerity and underlying changes in the patterns of demand, resulting from population growth, ageing and rising chronic disease.

**Evidence summary**

Evidence from the Organisation for Economic Co-operation and Development (OECD) and UK health care systems suggests that the way hospitals and other providers of health care are funded can have an impact on the quality and efficiency of health services (Busse and others, 2011). This evidence supports a role for the payment system in improving both quality and provider efficiency, but there is limited evidence to suggest it can be used to incentivise system efficiency or better patient outcomes (Street and others, 2011).

The last decade has seen rapid and radical reforms to the way both hospitals and primary care providers are paid in the NHS (Appleby and others, 2012). Payment reform has focused on two key areas:

- **How best to pay for units of health care?**
  
  The NHS introduced the Payment by Results tariff system of prospective, nationally fixed prices for case-mix adjusted units of acute care activity (for example, all services included in an inpatient stay for a hip replacement). Prior to this, block budget arrangements predominated, in which the level of funding for a hospital or other provider was locally negotiated and generally did not vary with the amount of health care provided within that year. In primary care, the NHS has a tradition of a significant element of capitation funding, in which the payment for core general medical services is bundled together and fixed for each patient for the year, weighted for measures of need and cost, but which does vary with the number of patients registered with the GP. Some registered patients will use more health care than this amount covers, while others will not access their GP within the year.
How best to incentivise high-quality care?
This often involves pay-for-performance in addition to the base payment for an activity or a weighted patient population. In primary care, the NHS introduced the Quality and Outcomes Framework (QOF) payment, in which a third of payments to GP practices are dependent on the achievement of quality and performance targets/standards. For hospital services, the NHS has piloted the Advancing Quality programme in the North West of England, and introduced best practice tariffs (BPTs) nationally. In addition, the Commissioning for Quality and Innovation (CQUIN) programme withholds 2.5 per cent of income from hospitals, mental health trusts and community health service providers, which they must ‘earn back’, conditional on achieving a mix of national and local quality and performance goals.

Payment by Results
The evaluation of the Payment by Results case-based prospective payment system for some hospital services found that the introduction of Payment by Results was associated with a more rapid reduction in unit costs and length of stay in NHS hospitals. In addition, there was a faster increase in the proportion of elective care provided on a day-case basis. Average length of stay fell for both planned and emergency care, although the impact was greater for planned care. The effects occur for most population groups with limited evidence of any distributional effects – older age groups experienced larger reductions in length of stay and younger age groups had larger increases in the probability of being treated as a day case.

There was little, if any, measureable change in the quality of care, suggesting that reductions in cost were achieved through improved productivity, rather than sacrifices in quality. This evidence is based on the first few years of the Payment by Results system (up to 2007/08), using hospitals in Scotland as a control and also comparing hospitals in England that introduced Payment by Results at different times.

The effect sizes are comparatively small. The introduction of Payment by Results was associated with a 0.5 per cent a year reduction in average length of stay and appears to have made a relatively small contribution to the NHS’s long-term reduction in length of stay and increase in day-case treatment. The reduction in average length of stay is equivalent to one bed day for every 23 spells of care provided and around a sixth of the reduction in length of stay which occurred between 2002/03 and 2007/08. This is a saving of between one to three per cent of resources to deliver the same amount of care (Farrar and others, 2007; 2009; 2010). Across Europe, the use of case-based payment systems similar to Payment by Results has also been associated with reductions in average length of stay.

The effect on the activity level of introducing case-based payments has depended on the preceding system. Where case-based payment systems have replaced block budgets – typically based on the historic cost of providing care – they have been associated with an increase in health care activity. Where they replaced fee-for-service payment systems, they have been found to reduce hospital activity (Street and others, 2011). The introduction of Payment by Results in the NHS was also associated with a statistically significant increase in the number of planned and emergency admissions (around nine per cent for
emergency care and three per cent for planned care between 2002/03 and 2007/08) (Farrar and others, 2010).

Although the evaluation results for Payment by Results are positive, it is important to note that there is no robust evidence on either the long-term impact of Payment by Results, or its impact on health system efficiency. The Payment by Results system has changed since its introduction a decade ago. Perhaps most significantly, the number of tariffs has increased substantially from 650 in 2003/04 to over 1,500 separate case-mix groups by 2012/13; prices for emergency activity are now paid at 30 per cent of the full tariff for all activity above the 2008/09 level; providers are not reimbursed for certain treatments if the patient is readmitted after a previous episode of care and so-called ‘never events’; and new bundles of payment are being developed which extend across more of the care pathway, such as the maternity pathway tariff and the cystic fibrosis tariff (Department of Health, 2012a).

There is also increasing evidence that the way the NHS is using the Payment by Results system has changed since 2007/08. Work by the Audit Commission found that for a number of years, payments to hospitals for services not covered by the national Payment by Results tariff were rising much faster than either the growth in activity for these services or payments for Payment by Results services (Audit Commission, 2011). Between 2007/08 and 2009/10, acute NHS trusts’ non-tariff income grew at over 20 per cent a year, while tariff income grew at less than three per cent a year. Subsequent work has found that 50 per cent or more of providers and commissioners vary the nationally set Payment by Results tariffs and rules through local negotiation (Monitor, 2012). In 2013/14, a survey of 91 non-specialist acute trusts found a net £2 billion of income shifting back to Payment by Results from local agreed payments, but with significant variation across the country. Just over half of all the acute trusts which provided data had seen a decrease in the use of off-tariff payment mechanisms such as block contracts in 2013/14, compared with 2012/13, while almost a third had experienced an increase in off-tariff payment mechanisms (Williams and Welikala, 2013).

In addition, as NHS funding growth has slowed, commissioners have become increasingly concerned about rising rates of hospital admissions. Standard contracting guidance encourages commissioners to agree on emergency activity limits in contracts, but until 2010, activity grew faster than in the period before Payment by Results was introduced (Audit Commission, 2011). The introduction of the national 30 per cent marginal rate tariff was an attempt to limit the incentive for increased emergency admissions, which is inherent in an activity-based payment system, and to share the financial risk of such increases between providers and commissioners.

In many areas, however, it appears commissioners have gone beyond this and have introduced additional arrangements for limiting commissioner exposure to activity risk. Some NHS organisations have agreed so called ‘cap and collar’ arrangements where the commissioners pay providers for activity using the national Payment by Results tariff, but only to a certain contract value, beyond which the commissioner will no longer pay (Healthcare Financial Management Association, 2013). Robust information on the nature and scale of such arrangements is not currently available, but will be for 2014/15, as in future all such local variations will need to be published and notified to Monitor and NHS England. Furthermore, some commissioners and providers have been working on innovative models of bundling acute care for some patient groups or conditions into
year-of-care or capitation payments spanning the spectrum of care, to support service integration. These innovative models shift more of the activity risk to a network or group of providers.

There is no evaluation evidence of the impact of these changes to the Payment by Results tariff, or to the way the NHS has implemented the tariff from 2007/08 onwards. However, the Swedish experience suggests that the productivity gains from payment reform may be transient or at risk of reversal. In Sweden, purchaser/provider split and activity-based funding were introduced in five counties in the late 1980s and early 1990s, during a period of budgetary contraction. Real public, per capita spending on health fell at a rate of about -1.2 per cent per annum in Sweden as a whole between 1989 and 1994 (OECD, 2003). This squeeze was associated with a rise in hospital efficiency in all counties (Tambour and Rehnberg, 1997). However, relative efficiency rose faster in the five counties that adopted activity-based funding than in the remaining 21 counties maintaining global budgets. The productivity effects seem to have been temporary, however, at least in Stockholm County (Mikkola and others, 2001), perhaps because volume ceilings were introduced (Kastberg and Siverbo, 2007) and perhaps because real public health spending began to expand again in 1994, rising at an annual rate of 2.8 per cent per annum between 1994/05 and 1997/08 in Sweden as a whole.

Pay for Performance

In many ways, the NHS has been at the forefront of developing and implementing pay-for-performance incentive schemes in both hospital and primary care settings. It has also commissioned a number of evaluations of the impact of these initiatives. Many of these evaluations have reported their findings in the last year, extending and deepening our understanding of the impact and limitations of the use of financial incentives to improve quality.

In primary care, the evidence suggests that the QOF pay-for-performance scheme has improved quality-related processes that and that the quality of care for chronic conditions has improved since its introduction (Maisey and others, 2008). In the early years of its implementation, the QOF may have contributed to reducing previously wide inequalities in primary care quality (Dixon and others, 2010). But attribution is difficult, given the range of policy initiatives introduced alongside the QOF, pre-existing trends towards higher quality, and the lack of a control group for comparison.

There is some concern that financial incentives may have had a detrimental impact on intrinsic motivation, specifically for practice nurses (Glasziou and others, 2012). There is also concern that single diseases have been emphasised in the QOF over multi-morbidity and there is relatively little focus on primary prevention, apart from smoking cessation (Dixon and others, 2010). Overall, there is no evidence that the QOF has had an impact on patient outcomes, but this may be as a result of the difficulties in evaluating this. Recent evidence suggests a potential saving of 11 lives per 100,000 people from the improvements in processes of care in the first year of implementation, but with no further
gains in the second year as practice performance typically exceeded the level that attracts maximum payment (Gillam, 2013). There is also some evidence of a reduction in the costs of hospital care (Health Foundation, 2011). Cost-effectiveness evaluation for some QOF indicators suggests the incentive payments were cost-effective, even with modest improvement in care, although not all indicators were cost-effective and no account was taken of the cost of administering the programme (Gillam, 2013).

The evidence for pay-for-performance schemes in secondary care is mixed. The Advancing Quality scheme was trialled in 2008 for 18 months in hospitals in the North West England Strategic Health Authority. It provided bonus payments for relative performance across 28 quality measures covering five clinical areas, including mortality and readmission rates. Advancing Quality introduced a wider range of quality improvement measures alongside the financial incentive. The results of the evaluation are some of the most positive in the research literature on pay-for-performance in health care, with reductions in mortality in three of the five incentivised conditions, estimated to be the equivalent of 890 fewer deaths over 18 months. Taking account of the cost of the bonuses alone, the evaluation found the Advancing Quality programme to be cost-effective (Sutton and others, 2011; 2012).

In contrast, the evaluation of the subsequent NHS-wide CQUIN programme found that CQUIN schemes were much more diverse. In 2010/11, there were over 5,000 indicators across 337 CQUIN schemes. Many indicators were focused on structures and processes rather than outcomes, and were based on evidence of effectiveness that was weak at best. While the CQUIN framework has been successful in helping commissioners and providers to jointly identify and prioritise local needs for quality improvement, the evaluation found that CQUIN did not lead to statistically significant improvements in outcome indicators, and the involvement of front-line clinicians was often lacking (McDonald and others, 2013).

Best practice tariffs (BPTs) were found to have more clinical support than CQUINs. The evaluation found that they were regarded as more evidence-based and fairer. The take-up and impact of BPTs varies across conditions. There was a low take-up for cataract surgery but much higher take-up for stroke, hip fracture and cholecystectomy. The evaluation found no impact on quality or outcomes for stroke over and above the improvements achieved nationally through other quality initiatives. For hip fracture, there is evidence of improvements in process quality and outcomes, and for cholecystectomy, improvements in the day-case rate (McDonald and others, 2012).

The evidence from the various pay-for-performance schemes introduced in the NHS across hospitals and primary care tells us that these schemes need to be carefully designed. Clinical support – which tends to be higher when programmes are evidence-based and where lead times are sufficient to permit clinical engagement – is critical. Pay-for-performance schemes are likely to be more effective if they are part of a package of quality improvement. Schemes may have an initial impact – accelerating improvement in the short term – but it is not clear how far pay-for-performance can sustain quality improvement.

Evaluation of the Premier Quality pilot hospital pay-for-performance scheme in the USA found that while quality improved faster in pay-for-performance sites initially, after five years, quality improvement in control sites converged with intervention sites (Jha and others, 2012). Complexity may also pose a risk to schemes having their intended effect, as the incentives are less clear; however, conversely, there is also a risk that if schemes are too
narrowly focused, gains in quality of incentivised metrics may be achieved at the expense of other areas being de-prioritised.

Looking across payment systems in different countries and different settings, the results of two systematic reviews suggest that, overall, financial incentives have generally been effective in improving processes of care such as referrals, admissions and prescribing costs. However, the extent of the evidence identified was limited as none of the studies examined the effects of financial incentives on patient outcomes. Looking specifically at primary care, systematic review evidence suggests payment incentive schemes had modest and variable effects on the quality of health care provided, with greater effects on process than outcome measures. However, many of the evaluations identified by these reviews were poorly designed and reported, leading to the conclusion that there is insufficient evidence as yet either to support or oppose the use of financial incentives to improve quality of primary health care. An added difficulty in drawing conclusions from studies set in different contexts is that the effects of different systems will vary depending on that context, and the prior system in place.

**Evidence conclusion**

The evidence indicates that payment systems can have an impact on the efficiency and quality of health care, but they are a blunt – and limited – tool. Furthermore, the impacts on quality of care and provider efficiency have been relatively modest and may be transient. All payment systems have advantages and disadvantages and there is no perfect system. Payment systems vary in the degree to which they bundle together different elements of care (fee-for-service payments being the least bundled and global capitation contracts for the continuum of care the most bundled). An effective payment strategy will need to combine different mechanisms, and be complemented by other levers for influencing quality and efficiency. The optimal mix will depend on the priorities of the NHS but should be informed by the strengths and limitations of the different systems, and the context to which they are being applied. The impact of the payment system on providers of care will depend on both the amount paid (the price) and the way care is packaged together into the unit to be paid for (the currency). Reform to the payment system needs to take account of both together. The impact of the payment system also depends on how contracting rules or negotiations determine the volume of units of care contracted for each year, how prices and volumes are adjusted from one year to the next, and any risk-sharing arrangements that apply over volume or cost changes during the year.

---

A key limitation to the design, implementation and success of payment systems is data

Changing the NHS payment system will take time. A key limitation to the design, implementation and success of payment systems is data – on costs, quality, and on process-outcome associations. There is an urgent need to address this, and to ensure approaches are grounded in evidence and target-desired objectives, rather than focusing on what is currently measurable. This will foster confidence in the payment system, which is currently lacking, resulting in reduced adherence and blunting of the intended incentives.
Recommendations for the longer-term development of the NHS payment system

Monitor and NHS England are now jointly responsible for developing the NHS payment system. They recognise that payment reform is an important policy tool (Monitor and NHS England, 2013c). As the evidence shows, payment systems can have a positive impact on quality and efficiency. But the evidence also points to the limitations of payment reform as a lever to drive organisational improvement – at least in large, complex organisations – and is lacking in respect of whole-system improvement. Provider payment mechanisms are not likely to be a lever for major service change. To ensure that the long-term reforms to the payment system support the NHS to deliver efficient, high-quality care, Monitor and NHS England need to develop an approach that has the following core features:

Comprehensive – reforms to the payment system have traditionally focused on a limited number of individual services, care settings or sectors. If the NHS is to improve the efficiency of the health system as a whole, it needs payment systems that cover the continuum of care and which create incentives for providing the ‘right care in the right setting’. Research commissioned by Monitor suggests that most of the possible productivity improvements available to the NHS over the next decade will come from the existing providers of NHS care delivering their services more efficiently (£5.6 to £10.3 billion) (Monitor 2013a; 2013b). As Table 1 shows, these productivity improvements are spread across all sectors of the NHS: acute hospitals, mental health and community services.

Table 1: Estimates of potential NHS provider productivity improvements (£ billion in 2010/11 prices)

<table>
<thead>
<tr>
<th>Service</th>
<th>Total spending in 2010/11</th>
<th>Potential productivity improvement – lower estimate</th>
<th>Potential productivity improvement – upper estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute care</td>
<td>£46.8bn</td>
<td>£2.7bn (6%)</td>
<td>£4.7bn (10%)</td>
</tr>
<tr>
<td>Primary care</td>
<td>£21.3bn</td>
<td>£1.2bn (6%)</td>
<td>£2.5bn (12%)</td>
</tr>
<tr>
<td>Community services</td>
<td>£8.4bn</td>
<td>£1.2bn (14%)</td>
<td>£1.8bn (21%)</td>
</tr>
<tr>
<td>Mental health services</td>
<td>£10.5bn</td>
<td>£0.5bn (5%)</td>
<td>£1.3bn (12%)</td>
</tr>
<tr>
<td>NHS</td>
<td>£87bn</td>
<td>£5.6bn (6%)</td>
<td>£10.3bn (12%)</td>
</tr>
</tbody>
</table>

Source: Monitor, 2013b.

But this work does point to some additional potential efficiency opportunities from better-integrated health and social care. Better-integrated care may also provide opportunities to improve quality, particularly for patients with chronic conditions and multi-morbidity. NHS England and Monitor need to develop a consistent and coherent approach to payment across all health services including community, mental health, emergency and elective hospital care and, crucially, primary care. This does not mean that the payment system should be the same for all these services, either in the approach to currency design or pricing, or in the use of pay-for-performance incentives, but it must be
consistent so that clinicians, provider organisations and commissioners are not faced with confused or competing incentives.

**Focused** – the payment system should be used to incentivise aspects of care for which there is evidence that payment systems are an effective lever (these being provider efficiency, transparency and accountability). Compared to many other countries, England has had a long and progressively increasing list of objectives for the payment system. The theoretical and evidence base for using the payment system to further these objectives varies considerably (Street and Maynard, 2007). The payment system must not be overloaded with unrealistic goals, but equally must not stand in the way of achieving wider health system ambitions. Equally, although there is much interest in linking payments to quality, or ultimately outcomes, the evidence base is limited and mixed – some pay-for-performance schemes show that financial incentives can influence quality, but not universally, and the impact may be transient and differ in different contexts.

**Aligned with wider system changes** – it is also important to recognise the limitations of the payment system as a lever for change. For example, changing the payment system alone is unlikely to deliver better integrated care, but an inappropriately designed payment system will act as a barrier to implementing more integrated models of care. As the evidence for pay-for-performance incentives shows, reforming the payment system alone cannot deliver the desired change. To be effective, the payment system must be part of a wider package of aligned incentives, regulation, data and information improvements, plus investment.

**Transparent and evidence-based** – incentive schemes can only work if the organisations and clinicians whose behaviour they are trying to change understand what is required. Too often, the incentives are blurred or inconsistent. In part, this is a result of the complexity of the current system. For example, by 2012/13 there were around 1,500 Healthcare Resource Groups (HRGs) in the Payment by Results system alone. The NHS has funded a considerable amount of evaluation research into payment system reforms. This is welcome, but there remain significant gaps and much has changed since Payment by Results was first introduced. Monitor and NHS England need to ensure that innovations in the payment system continue to be subjected to rigorous evaluation so the NHS can continue to learn about what works, and identify and act upon any evidence of unintended consequences.

**Credible and predictable** – delivering productivity improvements and service innovation takes time and often investment, so health care organisations need to be able to plan with greater clarity about the financial environment. Previously, prices have changed annually as reference costs are updated and currencies evolve. Since 2005/06, more than 40 per cent of prices have changed by ten per cent or more from one year to the next (with the exception of 2006/07 to 2007/08, when prices were not updated) (Monitor, 2012). Recognising this problem, Monitor has decided to keep relative prices broadly stable for 2014/15, using the 2013/14 prices as the basis for setting the tariff, rather than new reference cost data. Lack of credibility also results from the way payment systems are implemented by providers and commissioners. If non-tariff payments are adjusted to ‘compensate for’ or cross-subsidise gains and losses resulting from other aspects of the payment system, the incentives become at best opaque, and at worst undermined and irrelevant. This adds transaction costs to the NHS without realising the potential benefits of an effective payment system. But perhaps more importantly, it undermines the
integrity of the system, resulting in an NHS financial system characterised by ‘smoke and mirrors’, with the real drivers of the payment system hidden from view. In 2014/15, local commissioners will need to make public any local variations to the national system for both tariff and non-tariff payments. From this information, we should be able to better gauge the amount of variation and the nature of purpose, including:

- how much is focused on new currencies to support integrated care
- how much is risk management between commissioners and providers for activity pressures
- how much is *ad hoc* adjustment to enable providers and commissioners to break even when they cannot afford to comply with the payment system’s national rules.

Priorities for reform

Although many commentators argue that Payment by Results is no longer fit for purpose, given the challenges facing the NHS over the coming decade, nationally priced prospective payments for units or standardised bundles of care will, and should, continue to be at the heart of the payment system developed by NHS England and Monitor.

We argue this for a number of reasons. First, for as long as choice of provider is a feature of NHS policy, the system needs a payment system which supports patients exercising that choice, and encourages competition on quality rather than price. Research evidence suggests one of the key differences between the impact of the NHS patient choice reforms in 2006 and 2008, and earlier market reforms (the 1991 internal market), was the basis for competition. In the earlier period, competition was based on price, but following the 2006 reforms, Payment by Results fixed prices nationally and so competition focused on other differences between providers. A range of empirical studies – and literature from other countries – confirms that competition with fixed national prices can be associated with some positive quality effects (Bevan and Skellern, 2011).

The second reason for contending that nationally standardised metrics for units of activity will continue to be a key feature of the payment system is that, without these, it will be difficult to establish more innovative and workable payment models for wider bundles of services spanning the continuum of care. Also, without these core building blocks to benchmark service delivery costs and outcomes, where there are more innovative approaches, it will be difficult to hold providers and commissioners to account for the efficiency and quality of services. In many other countries that use case-based payment systems, the primary objective is to improve transparency and accountability for both resource utilisation and quality. Most countries in Europe use diagnosis-related group (DRG) systems that are broadly similar to the HRG classification that Payment by Results is based on. In some European countries (such as those without a commissioner/provider split), the DRGs are used as one among several metrics to calculate a fixed annual budget (other metrics are used, for example, to assess volume and mix of services required to meet population need) and the DRG is then used to monitor performance against the contract and used in the analysis of productivity and quality (Quentin and others, 2011).
Nationally standardised prices and metrics for activity are also important, where providers are trying to develop payment and contracting models that encompass the continuum of care across organisational boundaries – such as accountable care organisations (ACO). Examples of such models in the NHS include the work Oxfordshire CCG and the COBIC consortium have been leading to develop outcome-based commissioning for older people, mental health and maternity services across Oxfordshire (Oxfordshire CCG, 2013). These arrangements often involve capitation or year-of-care payments to prime contractors or alliances of providers. In such models, resource allocation and risk-sharing systems will need to be developed to share risk and return between the different providers within the ACO, whether they belong to the public, private or voluntary sector. They cannot effectively do this without measures of activity and unit cost. It would not be efficient in the NHS for every health economy of commissioners and providers to try to develop their own methodologies and systems for doing this. Already, commissioners seeking to pilot this type of contract face difficulties unbundling costs for particular pathways and patient groups from their block contracts for community health services. Moreover, allowing such variation may limit scope for subsequent competition within the market by raising entry costs.

This brings into focus the difference between the role of Monitor and NHS England as the national bodies with responsibility for designing the architecture of the payment system, and the role of commissioners and providers in implementing the payment system. Monitor and NHS England need to give providers and commissioners the tools to innovate in models of care and contracting arrangements, and the rules to ensure such innovation is in the best interests of patients and represents value for money. To do this, they need to create currencies that can be flexible, with the ability to bundle and unbundle services in different groupings, depending on local circumstances and evolving models of care, but always with transparency, to ensure that providers and commissioners can be benchmarked for the costs of services and the quality of care.

However, while recognising that case-based payments are likely to remain a core feature of the NHS for the foreseeable future, it is still the case that the payment system needs fundamental reform. This will take time, and we need to be realistic about the extent to which reforms to the structure of the payment system – as distinct from incremental changes such as reducing prices – can make a significant contribution to the current financial and productivity challenge.

Reforms to the payment system often have long lead times to put in place the data and systems to operate new approaches, and there is a further gap between changing the system and delivering an impact on the efficiency or quality of care. Policy-makers need to be realistic about how far payment reform can contribute to the financial and service challenges facing the NHS. That notwithstanding, we welcome the willingness of Monitor and NHS England to consider long-term and more systemic reform to the payment system. Drawing on the research evidence, we would suggest that Monitor and NHS England should prioritise longer-term reform of the NHS payment system which:

- fundamentally reforms the payment system for emergency care, shifting away from Payment by Results towards global budgets for capacity informed by standardised benchmarks
• develops nationally priced currencies for a wider range of non-emergency services, including community health services, to make it easier for CCGs to commission bundles or packages of care for pathways or patient groups spanning hospital and community-based services

• considers alternatives to funding the cost of capital which is currently part of the Payment by Results system.

These reforms will take time to design and time to implement. Their impact would probably not be felt for several years, even if there was significant commitment to reform and redesign of the payment system. In the shorter-term, our analysis suggests a number of areas where more rapid progress could be made:

• refocusing the pay-for-performance elements of the payment system towards improving the integration and coordination of care

• proactively engaging with the design and evaluation plans for payment pilots supporting integrated care and improved outcomes (such as capitation for frail older people in Cambridgeshire and Peterborough, or Bedfordshire’s year-of-care contract for musculoskeletal services), to maximise the potential to generate design and implementation lessons for the wider NHS

• improving the efficiency incentives in the payment system though significant extension of best practice tariffs to more planned care

• improving the evidence base to determine the efficiency factor to underpin pricing decisions.

Alongside these reforms, payment systems can only function effectively if they are based on accurate cost information, so there needs to be sustained focus on improving costing information across the NHS. Incentives can only operate if they are clear, consistent and transparent. Stabilizing the system so that the NHS has some certainty for planning and investment is also important.

Setting the efficiency factor and prices
Ensuring that the price paid for the service is set at the right level is of fundamental importance to the incentive effects of payment. This has often received less prominence in the debate about payment systems in the NHS. If prices are set too low, there is a real risk that quality of care will suffer, and that providers will not have an incentive to invest in sufficient capacity to meet demand; if prices are too high, commissioners will have to limit access or cut other services to balance their budget. Table 2 sets out the calculations used by the Department of Health previously, and now Monitor, to set the ‘efficiency factor’ which determines the change in unit prices paid to providers between one year and the next.
The evidence base on the scope for efficiency savings in NHS providers, realistic timescales over which change can be achieved, and the level of investment required to deliver such change, is in urgent need of significant improvement.

There is also an unresolved tension at the heart of the new NHS system. Commissioner budgets are fixed and are not growing in real terms until at least 2016. Previous research by the Nuffield Trust suggests savings of four per cent a year would be required for a decade if the spending review allocations for health continue to grow in line with inflation but no more (Roberts and others, 2012). The NHS Call to Action (NHS England, 2013b) confirms that this would result in a gap of up to £30 billion, or four per cent a year.

Prior to the current period of austerity, the Office for National Statistics estimates that between 1995 and 2010, UK-wide health service productivity grew at 0.4 per cent a year (Massey, 2010). The University of York finds that for the English NHS over more recent years (2006 to 2010), productivity grew at more than double that rate, at 0.9 per cent a year (Bojke and others, 2013). This is good news, but even if these trends continue, productivity growth will not fill the gap. These studies are necessarily backward looking. They cover periods of rapidly increasing resources when productivity growth may have been below its potential.

Looking forward, Monitor’s analysis of the potential productivity gains in the NHS (Monitor, 2013a) finds significant improvement opportunities of up to £18 billion over the decade (see Figure 1). But this is also insufficient to bridge the gap. Pay restraint has been an important source of savings but cannot continue indefinitely.

Based on current evidence, it will, therefore, be very difficult to argue for an efficiency factor of four per cent a year for very much longer. With continued efficiency requirements of this level, increasing numbers of providers will be pushed into deficit each year.
Reviewing the approach to funding capital as part of the payment system

As part of work on improving the price-setting element of the payment system, specific attention should be focused on the capital element of the cost of providing health care. Over the last few years, the NHS has significantly underspent its capital allocation. Capital spending in 2012/13 was £3.8 billion, compared to a capital budget allocation of £4.5 billion (Department of Health, 2013a). The Foundation Trust Network has raised concerns about whether the model for capital funding in the NHS is now working. The Payment by Results tariff currently includes the revenue costs of capital investment. At present, hospitals can only finance capital if they are able to build up surpluses. Surpluses can only be accrued if operating cost is below the tariff. Borrowing to finance capital is only possible if repayments on loans can be met from the difference between operating cost and tariff.

The current financial climate for the NHS makes servicing capital needs very difficult. Moreover, there may be more fundamental problems with this model. Organisations’ historical capital costs vary substantially in ways that are not entirely within the control of the current management of the organisation. Donated assets from charitable trusts are an important part of funding for some organisations, but access to such funds is patchy and not necessarily related to need. There are also issues for some hospitals with large private finance initiative (PFI) schemes which have debt servicing costs that consume a high proportion of these trusts’ operating expenses. Although small in number, these costs present a significant challenge to the individual organisations – this has been recognised and, in 2012, seven NHS trusts were approved for specific additional payments to help with their ongoing PFI costs, outside the standard payment system (Department of Health, 2012b).
Optimising capital stock is often integrally linked to service reconfiguration, with the very protracted and uncertain consultation and decision-making processes this involves. Monitor has commissioned work on the capital funding regime for NHS hospitals. This is welcome. The work needs to link with the payment reform agenda to inform whether the revenue costs of capital should continue to be part of the tariff system, or whether there is a more effective alternative. There are examples from other countries of capital costs being outside the tariff payment system, but this brings its own disadvantages in relation to incentives for optimising efficiency. The Audit Commission has previously questioned whether including capital in the payment system is the right approach – the tougher financial climate for the NHS makes this an even more pressing and pertinent question (Audit Commission, 2008).

Reforming the payment system for urgent and emergency care

While activity-based payment systems have advantages as payment tools for routine and planned care, it is not clear that the system of payment is the most effective and sustainable approach for emergency inpatient care, given the cost structures of these services and the limited role of provider competition for some elements of emergency care. Two-part tariff systems have previously been discussed as a refinement to the English NHS (Palmer, 2005; Audit Commission, 2008). A number of countries use global budget contracts for emergency hospital capacity, without risk-sharing. Increasingly, rich metrics and benchmarking tools are used to set target budgets and associated output and outcome targets.

Monitor and NHS England should consider moving away from national tariff prices for the emergency pathway for major emergencies, to a lighter form of regulation, based on local commissioners negotiating multi-year global budget contracts that build in incentives for providers to make efficiency gains and move towards key targets for service change (such as ‘right care, right place’ metrics). There are signs that we are reaching the limit of the ability of Monitor and NHS England to use the national tariff system to drive efficiencies in emergency care by bearing down on cost variation. There are indications that emergency inpatient care is being cross-subsidised from planned care and non-tariff revenue. A number of factors give rise to persistent variations in the cost of providing emergency care across trusts. In some locations, there is a need to balance access to care with efficiency. Research undertaken for Monitor on economies of scale and scope in the NHS suggests that elements of the emergency and urgent care pathway, such as Level 1 (major) A&E, have a very high proportion of fixed and semi-fixed costs, and the most efficient scale of unit is likely to exceed the size of the majority of A&E departments in the NHS (Frontier Economics and Boston Consulting Group, 2012). Transition costs of major service reconfigurations, poorly configured legacy estate, and configuration of other local services that play a role in emergency pathways (community health and social care, and primary care) are causes of variation that typically take some years to address.

Different populations tend to use different mixes of urgent care services for any given level of need, related, for example, to geographical factors (urban/rural) or level of deprivation (NHS England, 2013a).
A lighter-handed approach to regulation would allow commissioners and providers to tailor contracts to specific local service challenges and local causes of variation in efficiency. But it would still benefit from a national role for Monitor and NHS England in collection of nationally standardised data on activity, cost and other metrics for these services, in the development of benchmarking tools and in regulating for transparency. A deeper understanding of the potential productivity gains for different types of providers with different scale, configuration and patient characteristics should allow for a more tailored approach to setting efficiency targets for the emergency pathway that would reflect the potential of different providers for improvement. Monitor and NHS England should develop this benchmarking capability to strengthen yardstick competition in emergency care to promote efficiency. They should also consider whether fixed capacity payment models or two-part tariffs would be more appropriate for these services.

In developing such models, Monitor and NHS England would need to consider the range of emergency and urgent care services that could be included in a capacity payment system. In large part, the development of the payment system needs to follow the work on the future configuration of emergency and urgent care services, and in particular where patients are to be given a choice of services.

Competition exists for urgent care for minor conditions that do not require admission, with 111 and 999 services, pharmacies, urgent GP appointments, out-of-hours services, walk-in centres and A&E departments all providing these services, though with overlapping rather than fully substitutable service profiles. Models of care that are based on procuring urgent and the entire emergency care pathway as a bundle, to incentivise shifts of care away from hospital, will therefore involve a trade-off with patient-driven competition. This will also depend on the most efficient model of delivering care – while choice may be attractive to patients, competing services may actually add to cost and induce demand.

An added challenge is the relationship between hospital and community health service funding, which is covered by the statutory payment system, and funding for primary care, which is separate, mainly through the GP contract. As Figure 2 shows, key parts of the emergency care system are delivered by primary care, so paid specifically as urgent care (for example walk-in centres), but other elements are paid for from the core capitation system for GPs (such as urgent appointments at a GP practice).

Fundamental reform will take time. But in the interim NHS England could, for example, work on developing and encouraging the use of benchmarks/metrics for setting target volumes/mix of emergency admissions based on catchment population characteristics – as an alternative to current marginal rate rules for historic volume baseline plus marginal rate.

Extending the payment system to provide comprehensive coverage of hospital and community services

Monitor and NHS England should prioritise developing and mandating the collection of data on meaningful packages of care for community health services, linked to assessed patient need, along with much better data on the costs of these packages, mirroring the
programme already being implemented for mental health services. If commissioners and providers want to develop pathway or year-of-care capitated payments spanning community services as well as acute, primary and/or social care, this work would provide them with the tools for doing so. Alternatively, the packages of community health services could be commissioned separately, but as part of a coherent strategy for commissioning and coordinating care across provider boundaries.

**Figure 2: The future of the NHS payment system**

<table>
<thead>
<tr>
<th>Hospital and community health services</th>
</tr>
</thead>
</table>

- **Emergency inpatient care**
- **GP urgent care**
- **Walk-in centres**
- **Medical assessment unit**
- **A&E**
- **111**
- **999**
- **Maternity care**
- **Mental health services**
- **Specialist and tertiary care**
- **Outpatient care**
- **Diagnostics**
- **Recovery, rehabilitation and reablement pathway tariffs**
- **Community services**
- **Nationally fixed prospective case-based payments**
- **Provider-specific emergency payment (two-part tariff/capacity payment models)**
- **Primary care services**

**Refocusing the pay-for-performance elements of the payment system on improving the integration and coordination of care**

One of the key challenges facing the NHS is the need to support better integration of care across settings. There is considerable interest in the potential role that the payment system could play in supporting integrated care. Evidence in this area is very limited (Frontier Economics, 2012), but it will be important to be clear about the role payment reform can play in supporting better integration of care.

Monitor and NHS England also need to set out clearly the respective role of national and local innovative models of contracting and payment. In Figure 2 we suggest that NHS England and Monitor should focus on developing currencies and prices at a more
granular level and that wider bundling should be for local decision-making based on priorities and circumstances. However, there is one area where there may be scope for more national direction, and that is the pay-for-performance elements of the payment system. We suggest that Monitor and NHS England should explore whether at least part of the existing pay-for-performance schemes (QOF and CQUIN) and local enhanced service contracts could be combined and reformed to incentivise coordination of care across hospital, mental health, community and GP services, and more appropriate and efficient transitions of care between settings. In some ways, the requirement for GPs to be accountable leads for the care of older people is a step towards this, but it could go further by aligning the GP reforms with the CQUIN payments to other NHS providers delivering care to the elderly and other patients with complex needs.

There is growing NHS and international interest in global capitation payments systems covering the whole continuum of care for a defined patient population, as a reform to incentivise greater integration of care. New capitation models are more sophisticated than those used in managed care systems in the past, combining weighted capitation with pay-for-performance and risk-gain sharing (Frakt and Mayes, 2012). We think it is premature for NHS England and Monitor to make global capitation a key feature of their work on payment systems. This is for a number of reasons: first, without more detailed work on currency development and costing for a wider range of activity, it will be very difficult to develop robust provider capitation systems, particularly where they cross organisational boundaries, and difficult for capitated provider groups to allocate resource and share risk appropriately within the group. Second, the NHS has not yet developed the kind of regulatory regime that would be needed to protect patients from the risk of denial of care and quality reduction under capitation payment. Third, a major change in the payment regime like this would create windfall winners and losers, due to the kind of geographical variation in cost and utilisation noted above, at a time when the health system is facing increasingly constrained fiscal space for managing this kind of risk.

Improving the efficiency incentives for elective health care
Increasing the rewards for efficiency and outcomes improvement in payments for elective care should be a key focus of reform for Monitor and NHS England. The NHS has been extending the number and range of services covered by BPTs. In doing this, it is gradually moving away from paying for activity based on average historic costs, towards paying for efficient models of care. This is a step in the right direction and supports both the ongoing productivity challenge facing the NHS and also evidence-based clinical practice. We encourage Monitor and NHS England to be ambitious in the development of BPTs and to set out a three- to five-year plan for their extension.

Increasing the rewards for efficiency and outcomes improvement in payments for elective care should be a key focus of reform
As part of this, Monitor and NHS England need to review the approach to case classification. In particular, we encourage them to look towards opportunities to price whole episodes of care – from diagnosis to rehabilitation and reablement – and reward outcomes, as has been done in Sweden for hip and knee replacement, and surgery for back pain (Wohlin and others, 2011).

Monitor and NHS England also need to review what factors they take into account in determining unavoidable differences in the cost of providing care. The main focus on this has been the market forces factor. However, many in the NHS are concerned that the HRG system does not fully capture the cost of the most complex patients, particularly those with multiple co-morbidities. A recent study found that HRG codes explain around one-third of the variation in costs at the patient level, but that the relationship between patient-level costs and the level of adjustment provided by the market forces factor was particularly weak (Monitor, 2012). There are also questions about the impact of age and socioeconomic factors on the cost to the NHS of providing care. Implementation of the revised classification system, HRG4+, should begin to address these issues, but there is a need to develop a much more robust evidence base.

**Conclusion**

The way the NHS pays providers to deliver care matters. In recent years, the NHS has often been at the forefront of innovation in this area. Research examining the impact of these reforms provides new evidence to underpin further reform. The NHS faces new challenges and has a new regulatory framework. While the payment system is important, the evidence suggests that reform will take a long time and the effect of any changes is likely to be modest and highly dependent on other changes. There remains much to do in the short-term – improving the information base on cost and efficiency is essential, alongside encouraging and supporting innovation and evaluating its impact. But Monitor and NHS England have an opportunity to set a longer-term vision for the payment system and instigate work programmes to implement and evaluate fundamental reforms.

Figure 2 (page 21) sets out our thoughts for what such a system might be: at its core is a new approach to paying for urgent and emergency care that would move away from activity-based payment, to contracts for capacity, informed by agreed models of care such as those emerging from the Keogh Review (NHS England, 2013a) and standardised benchmarks. Alongside this, there would be a substantial extension of both best practice tariffs and nationally defined and costed currency units to provide comprehensive coverage of community and other health and social care services not currently covered by tariffs. This would be led by Monitor and NHS England, but with the aim of providing building blocks that could be combined locally by commissioners into new contract arrangements for wider packages of care such as year-of-care payments.
References


Department of Health (2012a) *A Simple Guide to Payment by Results*.


Department of Health (2013a) *Annual Report and Accounts 2012-13*.

Department of Health (2013b) *Payment by Results Guidance for 2013/14*.


Farrar S, Chalkley M, Yi D and Ma A (2010) *Payment by Results: Consequences for key outcomes measures and variations across HRGs, providers and patients*. HERU.


Frontier Economics and Boston Consulting Group (2012) ‘A study investigating the extent to which there are economies of scale and scope in healthcare markets and how these can be measured by Monitor’, Monitor.


Health Foundation (2011) *Do Quality Improvements in Primary Care Reduce Secondary Care Costs?*


Monitor (2013a) Closing the Funding Gap: How to get better value for patients.


Palmer K (2005) ‘NHS reform: will it deliver the desired health outcomes?’


About the authors

Anita Charlesworth
Anita leads the Nuffield Trust’s work on health care financing and market mechanisms. Before joining the Nuffield Trust as Chief Economist in September 2010, she was Chief Analyst and Chief Scientific Adviser at the Department of Culture, Media and Sport. Previously, Anita was Director of Public Spending at the Treasury, where she led the team working with Sir Derek Wanless on his reform of NHS funding in 2002. She has also worked as an Economic Adviser in the Department of Health and for SmithKline Beecham pharmaceuticals in the UK and USA. Anita is a non-executive director of Whittington NHS Trust and a trustee of Tommy’s, the baby charity.

Loraine Hawkins
Loraine is a Visiting Senior Fellow at The King’s Fund. She is a health consultant, specialising in health systems policy. Loraine worked as a Lead Health Specialist for the World Bank in Eastern Europe and East Asia from 1998 to 2006. Since then, she has worked on international health systems in developing and industrialised countries. Outside the World Bank, Loraine’s work has included secondments to the UK government’s Treasury and Department of Health, where she was a member of the strategy unit that developed policy in 2001 to 2002 on autonomous foundation hospitals and case-based-payment reform. She managed the project unit responsible for implementation of hospital payment reform in the Department of Health.

Louise Marshall
Louise Marshall is a Specialty Registrar in Public Health. As well as the Nuffield Trust, other public health training positions she has held include posts within the NHS (at primary care trusts and an acute trust), Public Health England and local government. Before starting specialty training, Louise worked as a research scientist at the Medical Research Council, evaluating primary care referral to a commercial slimming organisation, following a PhD in nutrition and obesity at the University of Cambridge.
For more information about the Nuffield Trust, including details of our latest research and analysis, please visit www.nuffieldtrust.org.uk