REAL Centre

Briefing: The future of the NHS hospital payment system in England

From recovery to transformation

July 2021 **Matthew Bell, Anita Charlesworth, Richard Lewis**

Key points

- How health care providers (such as hospitals) are paid is one way of influencing the quality and efficiency of NHS care. Over the coming years the NHS will face unprecedented challenges as it tries to recover the substantial backlog of unmet need after the pandemic, deliver on already demanding efficiency targets and redesign care, shifting services towards more community and primary care.
- Through the pandemic the NHS has been operating under an emergency payment approach to meet the exceptional costs associated with COVID-19. This approach will continue until at least October 2021. Over the coming months, policymakers need to decide what should replace these emergency arrangements.
- At the same time, the way the NHS is organised is changing. The new Health and Care Bill will repeal much of the 2012 Heath and Social Care Act provisions for competition and place increased emphasis on collaborative approaches. These will be led by new statutory bodies, integrated care systems (ICSs).
- Before the pandemic, the NHS payment system was moving towards a blended approach where providers receive an annual fixed payment supplemented by activity and quality-related funding – to replace the payment by results (PbR) tariff.
- With waiting lists at record highs, it might be tempting to move back to the PbR tariff to incentivise hospitals to treat many more patients. This would help address the backlog of care. But over the coming years the NHS also needs to improve care of an ageing

The Health Foundation

population with complex, long-term health problems. Coordinated care across hospitals, community, primary and mental health services is a priority and activity related payments are not well suited to this goal. Funding is going to be tight so improving efficiency and making sure the service does not overspend will also be key objectives. Even after the pandemic, a blended payment system to replace the PbR tariff remains the right direction of travel to help balance these potentially competing priorities for the NHS.

- While the broad direction of travel for the NHS payment system is clear, the devil is in the detail and COVID-19 has made determining the balance of fixed, activity and quality payments harder. Work needs to begin urgently on the specific design of a new strategy for the payment system if it is to be in place for 2022/23. Both to support the new ICS model and enable health services to recover from COVID-19.

Introduction

Directing the flow of funding through the NHS is one of the levers by which policymakers can try to influence the provision and cost of care. Evidence from the UK and further afield suggests that the way health care providers (such as hospital, mental health and community health services) are paid can influence the quality and efficiency of care.[1,2]

Over the coming years the NHS will face unprecedented challenges as it tries to recover the substantial backlog of elective care, deliver already demanding efficiency targets, and integrate care services – in part to facilitate shifting services out of hospitals and into community and primary care settings. In addition, ambitious long-term goals, such as better population health management and reducing health inequalities, are very unlikely to be met through the combination of the PbR tariff for hospitals and block contracts for community and mental health services.

This briefing considers these challenges and the role of a new payment system in meeting them as the NHS embarks on further reform and recovery from the pandemic. We begin by setting out the kinds of payment structures available to the NHS and describe the hallmarks of a 'good' system. We conclude by exploring what kind of system will be needed for the immediate COVID-19 recovery and to transform and meet the challenges of the post-pandemic environment.

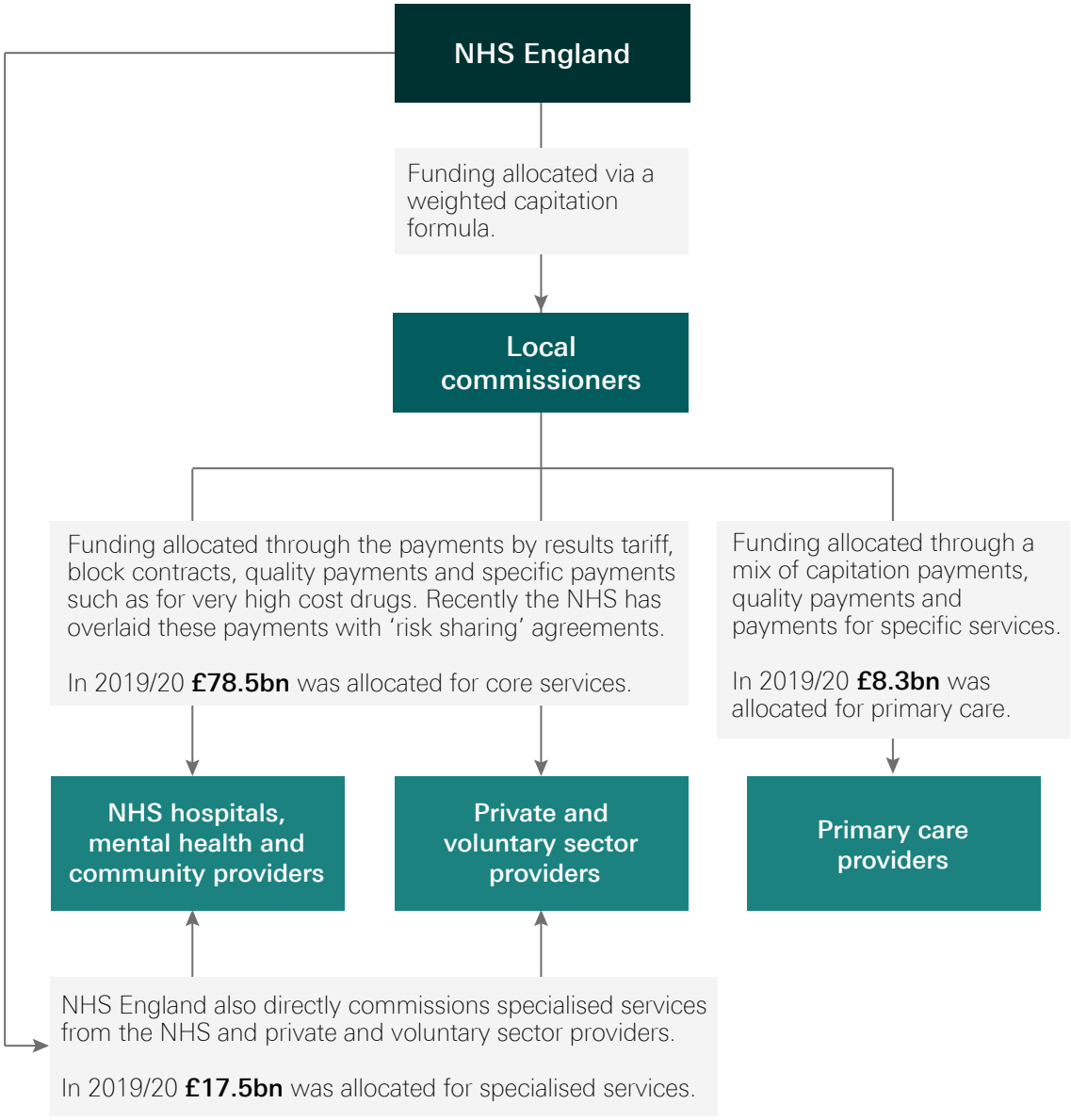
The briefing focuses on the payments for hospital, mental health and community health services rather than the payments to primary care providers (GPs, pharmacists, dentists and ophthalmologists) who are, in the main, independent contractors to the NHS.

What is the 'payment system'?

The payment system is an umbrella term for the arrangements by which funding flows from NHS commissioners of care to the providers, who deliver front-line care for patients.

NHS England commissions specialised services. Local health care commissioners oversee funding for local acute, mental health, community and primary care providers. Commissioners' budgets are determined by a resource allocation formula based on local population size, relative need and unavoidable cost differences (for example due to pay and building cost differentials in different locations). Under the Health and Care Bill, integrated care systems (ICSs) will take over this local commissioning role in England. Figure 1 sets out how funding flows to commissioners and onto providers.

Figure 1: How revenue funding flows from commissioners to care providers



Commissioners distribute funding to the providers of health care using a range of different payment approaches. These are set out in Box 1.

Box 1: Payment approaches by service area

The main payment method used to channel funding from the commissioners of care to NHS providers depends on the extent to which payments for services are paid for separately or grouped together – known as 'bundling'. The second difference is whether the payment rates are set in advance or retrospectively. During the pandemic, hospitals are being reimbursed retrospectively based on actual costs incurred due to the difficultly of anticipating an appropriate price to pay for key items.

Historically, the most common payment system in the NHS was an annual **block budget payment**. This provides an annual spending envelope for a hospital, regardless of the number of patients treated, the type of care provided or quality of care. Block contracts are still used extensively in mental health, community and ambulance services. The advantage of block budgets is that they provide certainty of income to providers and certainty of spending to commissioners. The disadvantages include limited incentives to improve efficiency or quality and to be responsive to patient needs.

The next most bundled approach are **capitation payments**. These are lump sum payments to cover the services an individual needs for a period of time – they are normally annual per person payments, often 'weighted' to take account of the fact that some patients require additional or more costly services. They can be used to pay for a year of care for someone with a specific disease (for example cystic fibrosis or diabetes) or for a range of care services such as those provided by GPs. Capitation payments are a way of incentivising providers to focus on better prevention and early intervention to avoid high cost later. They tend to work best when the cost of care for a group of patients with specific characteristics are reasonably predictable.

Activity payments have been rolled out across the NHS over the past 15 years through the Payment by Results (PbR) tariff system. In 2016 almost 90% of NHS contracts for acute health care were paid under the PbR system.[3] The PbR tariff provides a fixed payment for a specific episode of care (such as a hip replacement), taking into account the complexity of the patient's health care needs.[4] The two fundamental features of PbR are nationally determined currencies and tariffs. Currencies are the unit of health care for which a payment is made and are based on groupings of clinically similar diagnoses or procedures with a similar cost. Tariffs are the set prices for a given currency. Activity payments increase the incentive to provide services efficiently and encourage providers to deliver more care to more patients (assuming they are set at the right level). But the concern is that they can be a disincentive to collaboration, prevention and have comparatively high administrative costs.

The least bundled system of payment is **fee for service** where payment is made retrospectively for each unit of service provided; each activity or patient contact paid for according to a fixed price schedule, so rather than paying for a hip replacement as a whole, the commissioner would pay for the operation separately from any diagnosis tests and scans and outpatient appointments. Fee for service payment mechanisms encourage providers to be very responsive to patients' demand and deliver comprehensive (and, under some conditions, high quality) care but have weak incentives to improve efficiency, little incentive for prevention and often lead to over-treatment.

Elements of these payment structures can be combined to try to benefit from the strengths of each while avoiding some of their weaknesses. However, there is no 'perfect' payment system. In the end, the health service needs to match the payment system to its main objectives, recognising there are likely to be trade-offs between them.

In addition to these core payments there are top-up or 'withhold' payments for quality. In primary care this is the Quality and Outcomes Framework (QoF), which is a top up payment. For hospital, mental health and community services there is a system called Commissioning for Quality and Innovation (CQUIN), where in 2019/20 1.25% of funding was held contingent on providers evidencing good practice quality care in five areas.

In recent years the NHS has overlaid 'risk sharing' agreements on top of the PbR tariff. These agreements set out at the start of the year how much activity is expected to occur under PbR. They then agree what will happen if activity is either significantly higher or lower than planned. The agreements specify how the cost of that variation in activity would be shared between the

commissioner or provider. For example, if a hospital treats 5% more patients the risk sharing agreement might agree that the hospital should reimburse some of the PbR tariff back to the local commissioner to reduce the risk of the commissioner overspending. If the hospital treated 5% fewer patients than expected, the commissioner would agree to make an extra payment to help reduce the risk of the hospital going into deficit.

Service areas

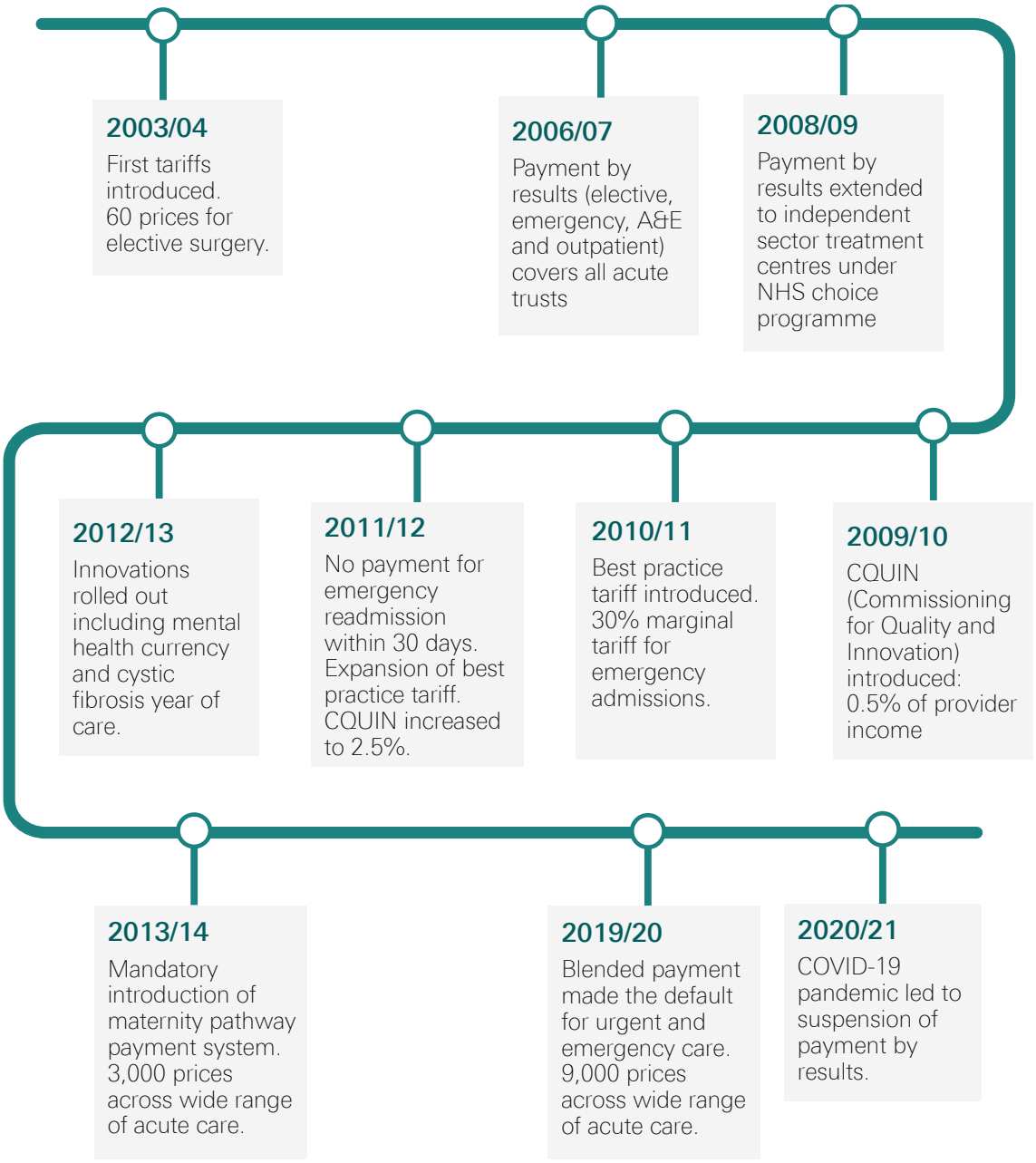
Mental health and community health: care providers' 'block contracts' predominate, where the total budget is agreed in advance with only indicative linkages to the amount and pattern of care. Some commissioners have used mental health care cluster data as a payment currency, with payment based on cluster days or episodes of care. A cluster day price is calculated by dividing the value of a provider's mental health contract by the number of expected days of care to be provided, weighted by the relative resource intensity of the different clusters. There are 21 clusters that provide a way of capturing the presenting needs of service users coming into mental health services. They group people with similar levels of needs in the same cluster, although their specific diagnosis may be different. The clustering tool has 18 scales (eg depressed mood, problems with activities of daily living). Each scale is given a rating from 0 (no problem) to 4 (severe to very severe problem). However, in 2015 NHS England and Improvement raised concerns[5] that payments based on cluster days would not be the best way to incentivise early intervention and recovery-focused care and was seeking to move to pay for mental health services based on either capitation or episodes of care.

Hospital care: until very recently there has been a move away from block budgets to activity-based payments based on the national average cost of providing an 'episode' of care. The PbR tariff has evolved over the past 15 years. Figure 2 shows the evolution of the PbR system up to 2019/20.

The shift away from PbR

The 2012 Health and Social Care Act emphasised the PbR system as a key pillar of a more market-based approach to the NHS, in which autonomous providers competed to deliver care. But the NHS is increasingly emphasising collaboration over competition, and seeking to incentivise integrated care.[6,7] A mismatch between policy intentions and payment incentives has emerged – and the ubiquity of PbR challenged. This section of the briefing explores why. Figure 2 provides an overview of changes to secondary care payments.

Figure 2: Timeline of key changes to the NHS secondary care payment system



In 2003/04 the NHS introduced a limited number of tariffs for elective surgery, for example £5,323 for a hip operation in 2012.[4] These tariff payments were designed to help to manage waiting lists during a period where reducing waiting times was a key priority and the NHS was receiving large annual increases in funding. Activity-based payments later grew in number and would cover a much wider range of services. The policy emphasis progressively changed, from incentivising additional elective care to reduce waiting times, towards 'efficiency' as NHS funding growth slowed dramatically after 2010. PbR sought to focus competition on quality, which is harder to observe.

During austerity the PbR tariff was used to set demanding efficiency targets for providers, direct competition was less important and there was more emphasis on 'yardstick competition'.* Increasing activity was no longer a goal, rather the aim was to limit the rate of growth in hospital activity. Head-to-head competition has increasingly been replaced by collaborative benchmarking and other tools to manage demand.

From 2016, with NHS deficits increasing, PbR was increasingly augmented with other forms of payment, such as risk-sharing and block contracts. By 2019, the amount of acute health care being paid for under the PbR system had fallen to just under 60% by value.[3]

The focus on 'efficiency'

The PbR tariff system focuses on a limited view of health care efficiency. The tariff incentivises lower unit costs and high quality for specific episodes of care. This is an important source of efficiency gains but narrow when taking a long-term, system-wide perspective. For example, in isolation, it fails to place enough weight on disease prevention, earlier diagnosis and integration across different types of care which reduces avoidable hospitalisations.

Some of the key opportunities to improve efficiency across the health care system will come from redesigning pathways to provide care in ways that avoid costly hospitalisations or late-stage treatment. The PbR tariff provides little incentive for the providers of care to focus on these opportunities.

Over time the payment system should also encourage rather than penalise innovation in medical treatments and processes for engaging with patients and the wider public. The PbR tariff system unintentionally made innovations more challenging, such as the move from face-to-face to telephone or digital consultations, as it took time to create new categories of services with an appropriate price that commissioners could use to reimburse the provider. COVID-19 has increased the rate of adoption of new technologies, for example transforming outpatient care with more appointments shifting to telephone or online. In a survey of NHS staff (total sample size 1,413) conducted by YouGov for the Health Foundation between 23 October and 1 November 2020, just over four-fifths of respondents (82%) said their organisation had increased its use of technology to some extent during the first phase of the COVID-19 pandemic (the period covering the beginning of the first lockdown in March 2020 to when the survey was conducted in October 2020).[8]

* Yardstick competition describes the process by which each hospital is forced to compete with its 'shadow hospital' based on nationally set prices, calculated from the average cost of delivering an episode of care adjusted for the expected rate of efficiency year on year across the NHS and any unavoidable differences in cost between hospitals. Yardstick competition is used in regulating monopolies where actual competition is not possible or desirable to incentivise efficiency. (Shleifer, A. 1985. A theory of yardstick competition. *The RAND Journal of Economics* 16(3) 319–327).

While there is much work to do to establish the most appropriate and inclusive use of digital services, it is clear that this will, and should be, part of the way care is provided in future.

These wider system considerations will be a key focus for the new ICSs and any future approach to payments needs to target a broader understanding of where the main opportunities to improve efficiency lie.

Emergency payment approach during the pandemic

In April 2020, COVID-19 led to the suspension of the PbR system entirely while the NHS reconfigured services to deal with the pandemic and as it faced significant new costs that were hard to precisely anticipate. In place of the PbR tariff, all trusts moved to a system of block contracts, with additional reimbursement if costs exceed the agreed amount. The value of the block payment was calculated nationally for each clinical commissioning group and provider based on recent past expenditure. Providers were then able to claim for additional costs to reflect genuine and reasonable additional marginal costs due to COVID-19'.[5]

As the NHS moves from the emergency phase of the pandemic towards recovery, the health service needs to develop a new payment system that supports the service to meet its challenges and priorities over the coming years.

In its February 2021 white paper, *Integration and innovation: working together to improve health and social care for all*, government outlined its intention to introduce new legislation to place ICSs on a statutory footing. The white paper proposes repealing much of the 2012 Act provisions on competition and amending key aspects of the legislation so that NHS England has more scope to reform the payment system. The Queen's Speech confirmed that the government would introduce a bill in the 2021/22 session of parliament to enact these changes and the Health and Care Bill was introduced to parliament in July 2021.[9]

Before considering how the payment system could be redesigned to meet the challenges of reform and recovery, we look at what constitutes an effective payment system.

What makes a good payment system?

Most payment systems need to balance competing objectives. First and foremost, the system for distributing money to care providers needs to ensure hospitals are reimbursed at a rate that reflects the unavoidable cost of delivering high-quality care and allows them to meet the needs of patients.

But the payment system also needs to incentivise wider objectives. There are financial goals – to maximise the efficiency of care and in a taxpayer-funded system, deliver financial control.[10] The payment system needs to play its part in ensuring providers and commissioners do not overspend and are financially

viable over the longer term. The payment system may also need to support wider system goals such as to reduce health inequalities, prioritise some areas of care (eg prevention) or to spur innovation.

The dimensions of efficiency

Many of these multiple objectives can be captured through various ways of looking at efficiency. Efficiency has several dimensions, all of which are important in the NHS:

- **Technical efficiency** – all else being equal, services are delivered at the lowest cost, for example by treating patients as day cases rather than admitting them overnight where evidence shows that clinical outcomes are the same.
- **Allocative efficiency** – resources flow to the most beneficial mix of services, for example through a better mix of prevention compared with treatment for the complications of diabetes.
- **Dynamic efficiency** – over time innovations can be introduced that improve the quality of care, lower its cost or both, for example the introduction of home monitoring technology to support people with long-term conditions.[11]

Balancing multiple objectives

Efficiency and financial control are not the only considerations. Other important objectives include delivering high-quality care, reducing inequalities and improving health outcomes. In addition to these national goals there may be other local objectives, such as supporting local economic development, reflecting the NHS's role as an 'anchor institution'.[12]

In 2017, Health Foundation research found that the NHS payment system lacked a clear overarching purpose, with NHS providers perceiving the existing system to have a multitude of objectives.[13] This is supported by international comparative research that identified 12 policy objectives for the current NHS payment system, in contrast with the three to five objectives pursued in other comparable European health systems.[2]

Multiple objectives give rise to the potential for tension and conflict. Different approaches to payment, such as block, capitation or episodic activity-based payments, are each better suited to achieving some objectives than others. Most research[1] suggests that for health care the optimal approach to payment is likely to involve a mix of approaches.

But balancing these different objectives through a mix of payment approaches can result in complexity. And with complexity comes the risk that payment incentives are opaque and not understood well enough to influence the behaviour of those receiving the payment in the intended way. For example, the number of tariffs

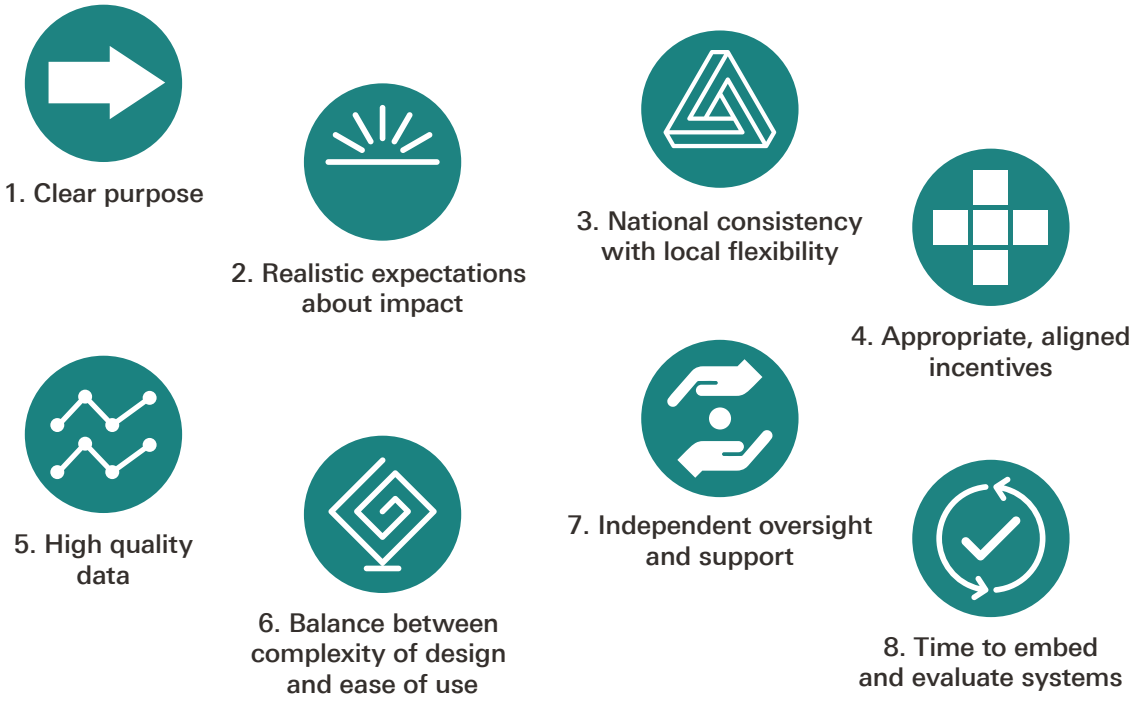
(many thousands) that eventually emerged under PbR reduced their influence on hospital decisions because hospital managers and clinicians found it difficult to keep track of the likely consequences of changes to treatment.

Complexity can be minimised by clarifying who is able to influence particular decisions. Financial risks created by the payment system should only be placed on institutions able to influence those risks. And the outcomes for which an institution is being rewarded (or sanctioned) must be, to a significant extent, under their control. For example, where a provider is rewarded for treating cancer but not the initial diagnosis, it is being held to account for the speed and effectiveness of its treatment but not for the severity of the cases being referred from primary care. If there is a view that the provider itself should be responsible for working with GPs to identify cases earlier on, then a different structure of payment is needed to reflect the ability of the provider to manage risk linked to treatment and to diagnosis.

Principles for the future payment system

Previous Health Foundation research[14] identified eight principles for future NHS payment systems (see Figure 3). These principles should inform the next steps in the design of the new approach to payments after the pandemic. COVID-19 will have an impact on NHS priorities but the principles outlined here are even more relevant to the payment system that needs to be put in place after the pandemic as they were before.

Figure 3: Eight principles for the future NHS payment system



Source: Wright et al. *Towards an effective NHS payment system: eight principles*. The Health Foundation; 2017 (https://www.health.org.uk/publications/towards-an-effective-nhs-payment-system-eight-principles).

NHS plans to reform the payment system

The *NHS Long Term Plan* set out three ambitions for the payment system, which were to:

1. take better account of the costs of delivering efficient services
2. ensure that the majority of funding is population based, and
3. align the payment system with the broader financial goals of the NHS.

Before COVID-19, the ambition was to move away from purely activity-based payments which reward the interests of individual organisations rather than systems. NHS system leaders wanted to harmonise the payment systems across different types of health care to aid population-based integrated care with greater collaboration between individual health care providers across a patient pathway.

The direction of payment reform is to replace the PbR tariff and block budgets with a blended system with three core components:

1. A fixed payment element to secure adequate supply of health care services. This would be set to cover the operational costs of providing the level of acute services agreed within population-level planning agreements.
2. A variable component to adjust payments to reflect differences between planned and actual levels of activity.
3. A quality component to incentivise and fund quality of care in priority areas.

In addition, the NHS proposed allowing local commissioners to agree risk sharing agreements to help ensure spending does not exceed the commissioner's funding allocation.

The impact of COVID-19 on the NHS

Even leaving aside the possibility of future waves of the virus, COVID-19 poses new and enduring challenges for the NHS that need to be considered alongside the goals set out in the *NHS Long Term Plan* – this raises the question of whether a blended system is still appropriate. As the NHS emerges from the pandemic it will need to tackle the backlog of unmet care need, meet new demands for mental health services and treating long COVID, and build resilience for future health shocks. COVID-19 will have major direct and indirect impacts on many dimensions of health service care:

- Demand: Lengthening waiting lists and waiting times:[15] the postponement of operations and diagnostic activities has created a growing backlog of cases. More than 300,000 people have been waiting more than a year for planned care (compared with around 1,500 in December 2019) and analysis suggests that waiting lists and median waiting times will grow for some time as the backlog of unmet need begins to present for care.

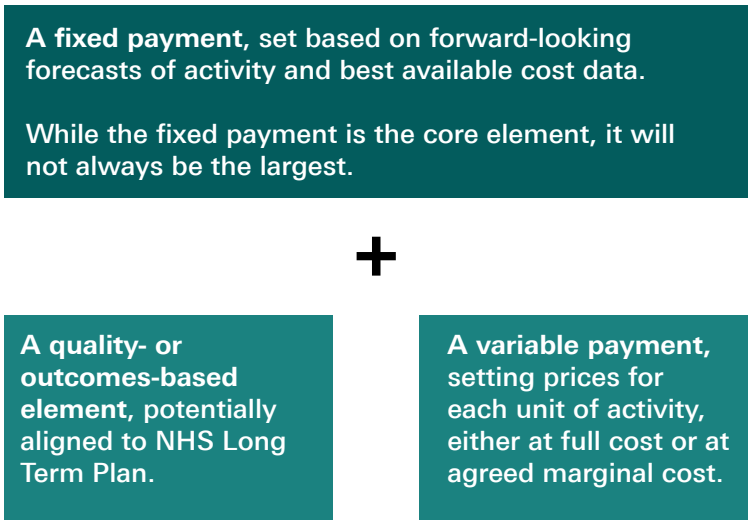
- Supply: Need for added capacity: the need for infection control requires new capacity to meet demand. Infection control measures have various impacts from the frequency of diagnostics to the use of operating theatres and wards. This will change the costs of delivering capacity such that historical data may not be a good guide to future costs.
- Demand: Rising mental health challenges: COVID-19 and its associated impacts (such as rising unemployment and social isolation) are leading to increased mental health burdens that will further add to waiting lists. Initial research[16] has documented rising levels of mental distress in the early stages of lockdown compared with pre-lockdown periods.
- Demand: Managing the longer term health consequences for patients who have had COVID-19: ONS research finds that in May 2021 around 1 million people self-reported living with long COVID, of which around a third first had COVID-19 at least a year previously.[17]
- Supply: the acceleration and use of new technology.

A blended system for the future

The complex challenges facing the NHS after the pandemic reinforce the need for a combination of payment approaches moving forward. No single approach is likely to be sufficient. For example, block contracts may reduce perverse incentives for providers to compete with each other for patients, but would do little to incentivise bearing down on the elective care backlog. While the scale of the backlog of care after the pandemic points to a need to incentivise additional activity, a move back to PbR alone for hospitals would not help the NHS meet its wider objectives. Activity-based payments do not provide a framework to fund reserves of capacity, which are necessary if the health system is to be more resilient to health shocks, such as another pandemic in the future. Nor do they address the poor coordination of care which may have a less visible impact than long waiting times but which is increasingly important given the number of people, particularly older people, who are living with multiple long-term health problems.

Figure 4 sets out the broad framework for a three-part blended payment system.

Figure 4: Framework for the hospital, community and mental health payment system



Source: Authors' representation of possible tariff system, drawing on NHSE and wider evidence base.

This three-part tariff framework has major advantages. However, there are some important underpinning design questions that need to be thought through if the new payment system is to support (and preferably stimulate) desirable service changes and outcomes without unintended negative consequences.

Moving towards a more system-based rather than individual provider organisation based approach to running the NHS, with a focus on population health gain, remains the longer-term direction of travel. But over the next few years the NHS is once again having to respond to a significant backlog of patients waiting for planned care and a need to increase hospital capacity. To address this challenge a larger share of provider funding will need to be channelled through the variable activity payment until the backlog can be cleared. The advantage of a blended payment is that as priorities shift over time so the balance of funding between the three elements can change. This provides a more adaptive system than traditional PbR or block.

Moreover, the public finances context means that the Treasury will be looking for the NHS to achieve high levels of annual productivity growth and ensuring the NHS does not move into deficit. While there is considerable agreement about the broad direction of travel for the payment system, the detailed design work for the new system will involve difficult choices and tensions, not least prioritising reduction of the elective care backlog in the short to medium term.

We now set out some of the issues that need to be considered in the detailed design work over the coming months.

Component one: A provider-specific fixed payment

The role of fixed payment

A core component of the blended payment system is a fixed payment, unique to the individual provider (hospital, mental health or community trust). The advantage of this approach is that it allows commissioners to set a payment that acknowledges the current financial reality of a provider and persistent variations in the cost of care, which may reflect the legacy of capital investments through the private finance initiatives (PFI) or lack of access to necessary capital to make services fit for purpose. It has been advocated for emergency care for a number of years[18] and was endorsed by the NHS for 2019/20

A key strength of this approach is that it allows local decision making about complex issues, such as balancing how far people have to travel to their health care facilities, which might require units that are smaller than optimal in terms of cost with efficiency.

Supporting the efficient delivery of services

It will be important to disentangle avoidable and unavoidable differences in costs between providers, so that organisations are not funded in perpetuity for inefficient care. If an ICS has a provider who is in deficit and has high costs, the ICS will need to set the fixed component based on a realistic assessment of the pace at which costs can be reduced and inefficiency tackled. But it is important that costs are no higher than absolutely necessary, and for no longer than necessary (ie providers need to be technically efficient). Conversely, a provider who can see how to reduce costs may wonder what benefit they would derive from doing so if the result is simply a lower fixed payment from the commissioner that could be used elsewhere.

To support these judgements, ICSs will need very good comparative data on costs and activity to develop efficiency benchmarks for services and whole hospitals, mental health and community providers. Comprehensive comparative data will be essential, and ICSs will also need support to interpret and use these data. A shift away from the national rules-based approach of PbR will be much more demanding for ICSs and the national role will be to support and develop ICSs.

As well as ensuring services are technically efficient, payments will need to reflect where best public value can be achieved ('allocative efficiency'). A fixed payment to hospitals could result in resources remaining in suboptimal settings (such as too much care remaining in hospitals rather than moving into home and community settings). The fixed payment element needs to be set at a level that reflects appropriate models of care.

Improving allocative efficiency may mean doing less of something that is not as cost effective in order to provide financial headroom for something more cost effective. Research has shown[19] significant differences in the cost effectiveness of spending at the margin in different disease areas (programme budget categories).

There is also evidence that within a given disease area there is scope for improved allocative efficiency. The NHS RightCare programme[20] identifies unwarranted variation in patterns of care and pathways across the NHS, and promotes tools for commissioners to shift towards more cost effective patterns of care.

Fixed payments will need to adjust over time if activity is moved into different care settings. However, local discretion over the timing of this will help with the management of 'stranded costs' that arise as pathways are changed. Stranded costs are those that cannot be reduced in the short to medium term despite activity having moved – these have historically proved an obstacle to service transformation.

Reflecting security of supply

COVID-19 has demonstrated the importance of 'security of supply'. In part this is about having sufficient capacity and flexibility within that capacity, which includes people, equipment (eg ventilators and scanners) and buildings. The payment system needs to reward the provision of that security of supply.

Security of supply is not only important in national or global emergencies, such as COVID-19. With the NHS regularly running at 90–95% capacity, there is too little margin to deal with any unexpected uptick in demand: a severe winter flu season, a significant heatwave or other events that may be more common than pandemics. All these risk stretching the NHS nationally or locally beyond its ability to effectively handle a crisis without jeopardising the health of those who need vital but non-crisis-related treatment.

The payment system needs to cover the cost of capacity that appears underutilised but is beneficial in terms of the system's resilience to external events (or unexpected variations in annual demand). The fixed provider specific payment will therefore need to cover planned NHS activity and an assessment of the unavoidable cost differences. But this also needs to reflect the system need for additional capacity, which will provide greater resilience to demand shocks from severe winters or new infections such as COVID-19.

Such payment structures need to balance the risk of failure due to insufficient standby capacity with the cost of building resilience. Working out how much extra capacity is optimal requires an analysis of risks. This is not a new concept: payments for security of supply are common in other vital services, such as electricity and water supply. Part of the 'fixed payments' received by these providers ensure there is sufficient capacity to meet unanticipated peaks in demand. Providers of these services are paid not just to supply current demand but to hold sufficient capacity for unexpected surges. In theory, unit prices such as PbR tariffs can be used to reward holding excess capacity: providers who hold excess capacity gain more revenue during the unexpected peaks. In practice, experience in other sectors suggests this revenue is too uncertain to lead to meaningful investment and that fixed payments are more effective.

Historically, in the health sector such 'excess' capacity has been viewed as wasted taxpayer money (beds empty, diagnostic equipment not used, staff trained for activities not carried out). A fundamental reassessment of this view is now essential given the NHS will be at capacity in the coming years and needs to build the resilience to deal with inevitable surges in demand, as well as the possibility of future shocks. Payments for security of supply should be incorporated into the fixed payment received by NHS institutions.

Scope of the fixed payment

A key question for the fixed payment element is whether this model is exclusively for NHS providers or whether ICSs should also use the same three-part tariff approach when contracting with non-NHS providers. If we need more health care capacity in the future for NHS funded care either to tackle waiting list backlogs and/or to provide greater resilience for health shocks, that capacity could all be provided in NHS owned hospitals or some of it in the charitable and independent sector.

The NHS drew on capacity from the charitable and independent sector in its response to COVID-19 and before the pandemic could use the 'any qualified provider' process to procure additional capacity (where, for example, waiting lists had grown too long). The government's white paper proposals aim to streamline this process, potentially still allowing the NHS to access wider capacity but without the need to use competitive processes by default. How far the payments for capacity and care provided by charitable and independent sector providers should match the system for NHS providers will be an important issue to work through. The NHS could continue to use the fixed tariff to pay private providers but the downside of this is that it does not deliver secure supply in the locations where it is needed.

Component two: Activity payments

Why continue with some activity payments?

To help reduce elective care waiting times, which were a significant problem in 2003/04, the PbR tariff was introduced.[21] The first decade after the tariff's introduction did see a rapid expansion in elective care and waiting times fell.[22] But the tariff was part of a package of reforms, funding and support that increased capacity and improved the management of patients on the waiting list.[23]

Even before COVID-19 the NHS was struggling to meet the 18-week waiting times standard[24] and the *NHS Long Term Plan* funding settlement included resources to improve access and reduce waiting times.[25] The pandemic has resulted in a major backlog in care. Modelling work projects that without a significant increase in activity, waiting times and the number of people on the waiting list will increase substantially.[26] From January to August 2020, 4.7 million[27] fewer patients were referred for elective care (compared with 2019). If three-quarters of these 'missing

patients' are referred for treatment in the coming months, the waiting list could grow to 9.7 million by 2023/24 compared with 4.4 million on the list before the pandemic in February 2020.

By the end of May 2021, nearly one-third (32.6%) of patients on the waiting list had waited more than 18 weeks. Meeting the 18-week standard by 2023/24 would require an annual average 11% increase in the number of elective procedures being performed each year by the NHS, needing some 4,000 extra consultants and 17,000 extra nurses a year. The Health Foundation estimates that it is more realistic to achieve the waiting time standard and eliminate the backlog of long waits over a 6-year period, ending in 2026/27. Achieving this would still require a significant increase in elective activity, but also additional funding of around £900m a year. In the 2020 Spending Review the NHS was allocated £1bn to begin to address the backlog of care but it will take many years of funding and additional capacity to get back to the standard of 92% of patients being treated within 18 weeks.

How to support and incentivise providers to increase elective care will therefore be a key issue once again for the NHS payment system over the rest of this parliament.

How important is the activity payment?

A key decision for policymakers and NHS system leaders will be how much of the overall payment system to devote to incentivising elective activity over the rest of this parliament. Alongside this is the question of how best to work with the charitable and independent sector. In 2004, the NHS introduced independent sector treatment centres (ISTCs) with multi-year contracts to try to increase capacity for tackling long waits. By 2008 as part of improving patient choice, the emphasis had shifted to the any qualified provider (AQP) approach where patients could choose where to have their care from the full range of NHS, licensed independent and voluntary sector providers with the PbR tariff ensuring money followed the patient's choice. This was associated with a big increase in the number of NHS-funded procedures undertaken in the independent sector.[28]

The white paper reaffirms that patient choice and AQP will remain part of the NHS 'offer'. The activity payment will need to align with the use of any qualified provider – ensuring fair payment to the new provider but also a level playing field with NHS providers, particularly in the balance between fixed and variable amounts.

The scale of the backlog of unmet need and the duration of the task to reduce it means that ICSs and providers will the NHS need to think very carefully about how to focus activity and capacity to prioritise patients in greatest need and deliver services efficiently. The payment system will then need to be aligned with these goals – channelling resources and focusing incentives on those elements of elective care that are prioritised.

Component three: Quality payments

Historically, the NHS has used the payment system to promote higher quality care through various combinations of best-practice tariffs, CQUIN payments and other targeted quality payments. Evidence suggests that at the right level, incentive payments for quality can be effective, particularly when supported by other measures to improve quality.[29]

It will be important to be clear about the objective for future quality payments. Historically, quality has primarily focused on measures of clinical effectiveness. The vision for population-based health shifts that focus to broader measures of wellbeing. Some of these measures align closely to the work of the NHS (eg obesity treatments, cancer diagnosis) but others extend beyond the NHS, incorporating complex and multi-agency interventions in local communities. This more recent development views the NHS as an anchor institution working in partnership with other public sector and civil society organisations to tackle broader health outcomes, inequalities and population-based metrics.

In the design of a new payment system, previous experience from CQUIN, QoF and best practice tariffs suggest that in designing the new payment system, the 'quality' or 'change' element needs to be sufficiently large to focus attention on these system goals. Our work with CEOs and senior managers suggests (albeit anecdotally) that they placed less focus on elements of CQUIN as its value fell over the past decade. Busy hospital managers and clinicians will not change established practice for small additional amounts of funding. Providers report that there is often a disconnect between the size of an incentive and the cost of achieving it.[14] The reduction of CQUIN from 2.5% to 1.25% of revenue by 2019 was, in our experience, accompanied by a similar reduction in its role driving change.

In contrast, QoF payments make up a much greater proportion of GP income. This meant that the GPs were more likely to take notice of the quality priorities included in the payment system (for example the proactive management of long-term conditions). And there is evidence that these performance payments to GPs accelerated quality improvements.[30]

Neither of these payments provides a perfect analogy (in part because the threat of withdrawing them was rarely seen as credible). But they contribute to a belief that 'size matters'.

The corollary of the need for a sufficiently large payment is that there has to be a process of prioritising where to focus in terms of improving quality. In which specific areas can the payment system really drive change if there is a clear signal of sufficient financial reward (and will there be unintended negative or perverse outcomes from providing this focus)? There is ongoing debate as to the extent to which these priorities should be set nationally or locally.

Wider aspects of an effective payment system

Previous Health Foundation research highlighted the key features of an effective payment system (see Figure 3). These encompassed several crosscutting areas that go beyond the design of fixed, activity and quality payments to include: high-quality data; independent oversight and support; national consistency with local flexibility, and evaluation. Here, we consider some of the key features that need to be addressed in the design of the post-pandemic payment system.

High-quality data

Designing a payment system without ensuring sufficiently accurate data are available will greatly reduce its effectiveness. A fixed payment between providers and commissioners will require good data on activity, case mix and the cost profile of the hospital. It may even require better (though not necessarily more) data than PbR about cost. The prices set in PbR are based on the national average cost of an episode of care, a provider specific fixed payment requires a more granular understanding of provider specific cost structures.

One important area of improvement has been the introduction of patient-level information and costing systems (PLICS),[31] mandated in the NHS from 2018. This programme will be critical to implementing a blended payment system, as will ensuring that PLICS data are timely. One challenge under the PbR system was that it was based on reference cost data that had long lags. A blended payment will require good comparative benchmarking data, for example the Getting It Right First Time (GIRFT) programme, which is using these data to examine unwarranted variations and improving efficiency and patient outcomes.

High quality data are important for all aspects of NHS service delivery, so the quality and use of data should be driven independently of the payment system. To work effectively the new ICSs will need access to high-quality data on population health needs, activity, quality and resources. ICSs will need an integrated data strategy and expertise to use and interpret that data. Transparency and accountability are also important. Data needs to be available to support good service delivery but also for public scrutiny. This will require skills and resources within each ICS but also national consistency and support – otherwise benchmarking is impossible and transaction costs may increase if providers have to give different information to each ICS.

Evaluation

For the payment system to succeed it will need to be continually evaluated, providing effective feedback to those involved on how well incentives are working. Unintended consequences or perverse incentives will also need to be identified quickly so the system can evolve and improve.

There needs to be a shared understanding of the expected outcomes, likely impact of changes to the system, and an understanding of how long it might take for these changes to become effective. It may take time for changes to the system to translate

into improvements for patients. For instance, a recent study into an integrated care transformation programme[32] shows that despite an initial increase of hospital activity, it took 6 years for the integrated care initiative to deliver a reduction of both A&E attendances and emergency admissions.

It is important the health system and evaluators make the most of data that are routinely collected in the health service, and combine this quantitative evidence with local qualitative insights to understand what is driving the change, and allow for course correction early on.

Independent oversight and national/local balance

Changes to the system will take time to design and will not implemented all at once. Moreover, the precise design will change over time as what works becomes clear. This raises the question of who should be in charge of this shift.

The proposed NHS legislation outlined in the government white paper will overhaul the national governance of the payment system. In the 2012 Act, NHS England had responsibility for the structure of the payment system (eg definition of currencies – *what* was being paid for) and Monitor (now NHS Improvement) for the level of the individual prices within that system (*how much* was to be paid). Meanwhile, a process of consultation and approvals for modifications was mandated with the Competition and Markets Authority (CMA) in a backstop regulatory role.

The proposed legislation merges the national bodies, abolishes the CMA role and gives NHS England greater flexibility as it reforms the payment system. The bill will introduce a new NHS payment scheme to replace the NHS tariff. Powers are proposed for NHS England to publish the payments scheme, and to give NHS England a power of direction over commissioners where they fail to comply with rules in the payment scheme. A power is also created for the secretary of state to prescribe, through regulations, some aspects of the proposed NHS payment scheme.[33]

Two distinct issues are important: who makes the initial decisions and what independent oversight exists to act as a check on those decisions?

Who makes the decisions?

A key question for the new payment system is where the balance of decision making will lie. The creation of large ICS collaborative arrangements could provide an opportunity for national policymakers to delegate more tariff-setting powers. The creation of hospital provider collaboratives within ICSs also raises the prospect that the right parties will sit at the same table and share mutual accountability to get the design right (and will allow representatives from other sectors to take part in this important discussion too).

However, the degree of autonomy given to local systems to determine the precise form and size of each of the elements will also be based on how much national variation is acceptable and the advantages and disadvantages of such variation. Elements such as security of supply might best be developed at reasonably large

levels of total population. Other elements (eg the balance of activity and fixed payments) might vary depending on how well waiting lists or chronic conditions (such as depression or obesity) are being handled in different parts of the country, unless there is a national programme.

The devolution of payment decisions from national bodies to local commissioners was underway pre-COVID-19, but this next step will require a more structured approach to such devolution.

Independent oversight of decisions

Devolution also raises questions about transparency and accountability. It is particularly important to consider how such a system would work in cases where there are significant financial challenges. The tensions between the financial stability and the sustainability of existing hospitals, quality of care and the need to shift funding to primary, community and mental health settings are hard to resolve and are best served by a transparent system where trade-offs can be considered. The 2012 Act sought to limit price competition due to concerns that quality of care might suffer. This is because quality is harder to observe and if providers and commissioners were under financial pressure they might set prices too low to maintain an appropriate quality of care. The inclusion of quality payments within the blended payment system is designed to ensure that quality remains a key focus but independent oversight of decisions to ensure prices are not being set below the cost of delivering high quality care will be important.

It is not clear where the independent oversight will come from in the new regime envisioned by the white paper and the Health and Care Bill. Parliament can provide some scrutiny but cannot provide the detailed checks-and-balances that institutions such as independent regulators or the CMA play in other parts of the economy. The National Audit Office or others may play that role. It will be important to identify where oversight sits as part of the ongoing reforms.

Conclusion: a payment system for the recovery and transformation

The whole health service – indeed the whole country and our economy – needs to recover well from COVID-19 and transform to meet the challenges of the post-pandemic environment. A sustainable health system needs to use public funds efficiently, adapt and evolve to meet changing patient needs and respond to public expectations.

A new payment system can play a part in supporting such a transformation. Incentives in the payment system must be designed to encourage all parties to work towards the same or aligned objectives. Shared objectives will foster effective relationships between providers and the new ICSs. Achieving goals, such as

population health management and reducing health inequalities, are very unlikely to be met through old methods such as the PbR tariff and block budgets. On this point, the NHS has reached a consensus.

A single approach to a payment mechanism, and the incentives contained within, will also not address the range of complex challenges facing the NHS over the next decade. Therefore, a blended approach seems most appropriate. However, identifying the precise nature of this blend is challenging.

We argue that the system being considered by the NHS before the pandemic based on three elements – a fixed payment with additional activity and quality payments – remains fundamentally right. If designed well, a three-part payment system could help to ensure the right capacity is in the right place, support care transformation with appropriate resource shifts, enable a reduction in the backlog of elective caseloads and ensure that the focus on activity does not drive out a focus on quality and population health. Local discretion in setting elements of the blended system would allow the weight on particular elements to be varied based on local needs. But that discretion needs to be accompanied by transparency and accountability and a national led evaluation programme that provides rapid learning to guide the evaluation of the payment system will be critical.

The scale of the challenges facing the NHS are daunting. To some extent, the NHS is simultaneously facing the challenges of the early to mid-2000s (waiting and capacity), alongside those of the 2010s (efficiency and financial control) and the 2020s challenge of population-based health improvement. While there is considerable agreement about the broad direction of travel for the payment system, the detailed design work for the new system will involve difficult choices and tensions as the service balances these, sometimes competing challenges.

If the health service is to have a payment system that effectively supports wider system goals it will be very important to recognise there will be an element of experimentation as new approaches are rolled out. These should be evaluated – in as close to real time as possible – to allow improvements to be made.

Payment systems are complex and not the most obviously engaging aspect of health care policy – but they do matter. They are the wiring that gets the right amount of money to the right part of the system to ensure needs are met with high quality, efficient and effective care. When the wiring is working, nobody notices. When the wiring fails, problems are felt throughout the health system. Getting payment reform right needs to be high up the agenda for the incoming chief executive of the NHS.

References

1. Marshall L, Charlesworth A, Hurst J. *The NHS payment system: evolving policy and emerging evidence*. Nuffield Trust; 2014 (https://www.nuffieldtrust.org.uk/files/2017-01/2014-nhs-payment-research-report-web-final.pdf).
2. Busse R, Geissler A, Quentin W, Wiley M. *Diagnosis-related groups in Europe*. Open University Press; 2011 (https://www.euro.who.int/en/publications/abstracts/diagnosis-related-groups-in-europe-2011).
3. Westminster Health Forum. *Payment systems and reimbursement in healthcare*; 2020 (https://www.westminsterforumprojects.co.uk/publication/payment-systems-and-reimbursement-healthcare-2020).
4. Department of Health. *A simple guide to Payment by Results*. Department of Health; 2012 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf).
5. NHS England. Consultation on payment proposals for mental health services for adults and older people commissioned by CCGs in 2016/17 [letter] 21 October 2015; (https://www.england.nhs.uk/wp-content/uploads/2015/10/mh-consult-letter-oct15.pdf).
6. NHS England. *The NHS Long Term Plan*. NHS England; 2019 (https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/).
7. Lewis, R. More Reform of the English National Health Service: From Competition Back to Planning? Int J of Health Services; 2018: 1, 5-16 (https://doi.org/10.1177/0020731418797977).
8. Horton T, Hardie T, Mahadeva S, Warburton W. *Securing a positive health care technology legacy from COVID-19*. The Health Foundation; 2021 (https://www.health.org.uk/publications/long-reads/securing-a-positive-health-care-technology-legacy-from-covid-19).
9. UK government. Health and Care Bill (https://bills.parliament.uk/bills/3022).
10. Lafond S, Charlesworth A. *A perfect storm: an impossible climate for NHS providers' finances? An analysis of NHS finances and factors associated with financial performance*. The Health Foundation; 2016 (https://www.health.org.uk/publications/a-perfect-storm-an-impossible-climate-for-nhs-providers%E2%80%99-finances).
11. Michalos, A C. *Encyclopaedia of Quality of Life and Well-Being Research*. Springer, Dordrecht; 2014 (https://doi.org/10.1007/978-94-007-0753-5).
12. Allen M, Malhotra A M, Wood S, Allwood D. *Anchors in a storm: Lessons from anchor action during COVID-19*. The Health Foundation; 2021 (https://www.health.org.uk/publications/long-reads/anchors-in-a-storm).
13. The Health Foundation. *Queen's speech shows government recognises major health challenges facing the country but not the scale of action needed. Health Foundation response to the Queen's speech 2021*; 2021 (https://www.health.org.uk/news-and-comment/news/queens-speech-shows-government-recognises-major-health-challenges-but-not-scale).
14. Wright A, Cornick E, Jones H, Marshall L, Roberts A. *Towards an effective NHS payment system: eight principles*. The Health Foundation; 2017 (https://www.health.org.uk/publications/towards-an-effective-nhs-payment-system-eight-principles).
15. NHS England. Consultant-led Referral to Treatment Waiting Times Data 2020-21 [webpage]. *NHS England* (https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/rtt-data-2020-21/).
16. Pierce M, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *The Lancet Psychiatry*. October 2020; 7, 10: 883-892 (https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(20)30308-4/fulltext).
17. ONS. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 4 June 2021 [webpage]. *ONS* (https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/4june2021).
18. Charlesworth A, Hawkins L, Marshall L. *NHS payment reform: lessons from the past and directions for the future*. Nuffield Trust; 2014 (https://www.nuffieldtrust.org.uk/files/2017-01/2014-payment-reform-policy-response-web-final.pdf).
19. Spackman E, Hinde S. *Methods for estimation of the NICE cost-effectiveness threshold*. University of York; 2011 (https://www.york.ac.uk/che/seminars/economic-evaluation/2011/methods-for-estimation/).
20. NHS England. NHS RightCare [webpage]. *NHS England* (https://www.england.nhs.uk/rightcare/).
21. Harrison A, Appleby J. Reducing waiting times for hospital treatment: lessons from the English NHS. *Journal of Health Services Research & Policy*. July 2009; 14, 3: 168-173 (https://www.jstor.org/stable/26751026).
22. Tallack C, Charlesworth A, Kelly E, McConkey R, Rocks S. *The Bigger Picture. Learning from two decades of changing NHS care in England*. The Health Foundation; 2020 (https://doi.org/10.37829/HF-2020-RC10).
23. Timmins N. *The most expensive breakfast in history. Revisiting the Wanless review 20 years on*. The Health Foundation; 2021 (https://doi.org/10.37829/HF-2021-C05).
24. Charlesworth A, Watt T, Gardner T. *Returning NHS waiting times to 18 weeks for routine treatment. The scale of the challenge pre-COVID-19*. The Health Foundation; 2020 (https://www.health.org.uk/publications/long-reads/returning-nhs-waiting-times-to-18-weeks).
25. Charlesworth A, Watt T, Gershlick B, Firth Z, Kraindler J. *Investing in The NHS long term plan. Job done?* The Health Foundation; 2019 (https://www.health.org.uk/publications/reports/investing-in-the-nhs-long-term-plan).

26. Kraindler J, Rocks S, Charlesworth A, Tallack C, Barclay C, Idriss O, Shembavnekar N. *Spending Review 2020: Managing uncertainty*. COVID-19 and the NHS long term plan; 2020 (https://www.health.org.uk/publications/long-reads/managing-uncertainty).

27. Gardner T, Fraser C, Peytrignet S. *Elective care in England Assessing the impact of COVID-19 in 2020 and where next*. The Health Foundation; 2020 (https://www.health.org.uk/publications/long-reads/elective-care-in-england-assessing-the-impact-of-covid-19-and-where-next).

28. Arora S, Charlesworth A, Kelly E, Stoye G. *Public payment and private provision. The changing landscape of health care in the 2000s*. Nuffield Trust and IFS; 2013 (https://www.nuffieldtrust.org.uk/files/2017-01/public-payment-and-private-provision-web-final.pdf).

29. Meacock R, et al. Paying for improvements in quality: recent experience in the NHS in England. *Nordic Journal of Health Economics*. 2014 2, 4 (https://doi.org/10.5617/njhe.794).

30. Roland M. Quality and Outcomes Framework: what have we learnt? *BMJ*. 2016; 354:i4060 (https://doi.org/10.1136/bmj.i4060).

31. Harrison K. *The benefits of exploring the costs of care on a case by case basis*. Imperial College Health Partners; 2019 (https://imperialcollegehealthpartners.com/the-benefits-of-exploring-the-costs-of-care-on-a-case-by-case-basis/).

32. Clarke G, Pariza P, Wolters A. *The long-term impacts of new care models on hospital use. An evaluation of the Integrated Care Transformation Programme in Mid-Nottinghamshire*. The Health Foundation; 2020 (https://www.health.org.uk/publications/reports/the-long-term-impacts-of-new-care-models-on-hospital-use-midnotts).

33. DHSC. *Health and Care Bill: delegated powers memorandum*. DHSC; 2021 (https://www.gov.uk/government/publications/health-and-care-bill-delegated-powers-memorandum).

About the authors

Matthew Bell is a Director at Frontier Economics where he leads Frontier's Health and Care work.

Richard Lewis is an independent consultant and Senior Associate at Frontier Economics where he advises on health care reform.

Anita Charlesworth is Director of Research and the REAL Centre at the Health Foundation.

About the REAL Centre

The Health Foundation's REAL Centre (research and economic analysis for the long term) provides independent analysis and research to support better long-term decision making in health and social care.

Its aim is to help health and social care leaders and policymakers look beyond the short term to understand the implications of their funding and resourcing decisions over the next 10–15 years. The Centre will work in partnership with leading experts and academics to research and model the future demand for care, and the workforce and other resources needed to respond.

The Centre supports the Health Foundation's aim to create a more sustainable health and care system that better meets people's needs now and in the future.

Acknowledgements

The authors are grateful to those who commented on previous versions of this briefing. Errors or omissions remain the responsibility of the authors alone.

When referencing this publication please use the following URL: https://doi.org/10.37829/HF-2021-RC17

About the Health Foundation

The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK.

Our aim is a healthier population, supported by high quality health care that can be equitably accessed. We learn what works to make people's lives healthier and improve the health care system. From giving grants to those working at the front line to carrying out research and policy analysis, we shine a light on how to make successful change happen.

We make links between the knowledge we gain from working with those delivering health and health care and our research and analysis. Our aspiration is to create a virtuous circle, using what we know works on the ground to inform effective policymaking and vice versa.

We believe good health and health care are key to a flourishing society. Through sharing what we learn, collaborating with others and building people's skills and knowledge, we aim to make a difference and contribute to a healthier population.

The Health Foundation
8 Salisbury Square, London EC4Y 8AP
+44 (0)20 7257 8000
e info@health.org.uk
@HealthFdn
www.health.org.uk

ISBN: 978-1-911615-62-0
Registered charity number: 286967
Registered company number: 1714937