

The NHS payment system: evolving policy and emerging evidence

Research report

Louise Marshall, Anita Charlesworth and Jeremy Hurst

February 2014

About this work programme

The health service has historically been a leader among public services in embracing innovation in payment systems. A wide range of different approaches to payment are now in place for different sectors and different areas, often introduced specifically to drive quality, volume or productivity improvements.

Our experiences with these systems are a source of lessons about when, and how, changing the way we pay for care can help achieve positive outcomes. Monitor and NHS England, now responsible for overseeing the payment system, have made clear that they see reform in this area as a significant part of supporting the system to improve quality and efficiency.

The Nuffield Trust has looked to support policy-makers by analysing the evidence for current systems and new options. We held a summit in 2012 looking at how health systems across Europe are attempting to use payment mechanisms to drive change, and compiled a report which drew out cross-country comparisons and promising initiatives from across the continent. In 2013, we published a case study looking at the possibilities for hospitals to use patient-level information and costing systems to improve efficiency.

This research report takes a comprehensive look at different approaches to payment and how they function in the English NHS. It then brings together and summarises the available evidence on whether recent payment initiatives have met their goals. It is published alongside a policy response, *NHS payment reform: lessons from the past and directions for the future*, which gives key conclusions on what is possible through short- and long-term payment reform, and looks at how we can achieve the best results.



Contents

List of figures, tables and boxes	2
Key points	3
1. Introduction	5
Approaches to paying for health care	6
Design of a payment system for health care	9
Objectives for a reformed payment system for NHS care in England	10
2. Payment systems in the English NHS	12
Who pays for what?	12
Evolution of payment systems in the English NHS	14
Taking a wider view	21
3. Review of evidence from evaluation of NHS payment programmes	23
Primary care Quality and Outcomes Framework	24
Activity-based payment in secondary care	25
Pay-for-performance schemes in secondary care	30
Block contracts in secondary care	33
Summary of evidence	33
4. Summary	34
References	36

List of figures, tables and boxes

Figures

Figure 2.1: Health care purchased by PCTs in 2012/13	12
Figure 2.2: The commissioning landscape in England since 1 April 2013	13
Figure 2.3: Funding GP practices in England	15
Figure 2.4: The evolution of Payment by Results in the NHS	17

Tables

Table 2.1: The components of the annual change in Payment by Result prices	18
Table 3.1: Hospital activity and length of stay under diagnosis-related payments	29
Table 3.2: Policy objectives for DRG payments in European countries	30

Boxes

Box 1.1: Payment methods in health care	7–9
Box 1.2: Key considerations when designing pay-for-performance schemes	11
Box 2.1: Pay-for-performance schemes in NHS hospitals	19–20
Box 3.1: International evidence on diagnosis-related payment	28

Key points

The current NHS payment system has evolved greatly over the last decade and employs a blend of different payment methods across different services and sectors. But the payment system for NHS-funded services is in need of reform to support a health system that meets the needs of an ageing population, with growing prevalence of multiple chronic diseases, in the context of tightly constrained resources.

Moves away from block budgets to activity-based payment approaches have improved provider productivity. However, while improvements have concentrated on acute care through the introduction of activity-based payment, block contracts are still the predominant payment mechanism for community sectors. Moreover, the structure of incentives across services does little to support policy ambitions to shift care out of the hospital setting, with the payment systems often giving conflicting signals. The predominance of activity-based payment in the acute sector, introduced at a time of long waiting lists, encourages activity in hospitals; at the same time, block budgets in community services and capitated budgets in primary care offer little incentive to increase activity or efficiency in these settings.

The shift from block budgets to activity-based payment in the acute sector has supported patient and commissioner choice, making providers compete on the basis of quality rather than price.

However, prices paid are now being reduced in an attempt to promote further provider productivity improvements. If the prices are insufficient to cover cost, there will be a risk to quality. To mitigate against this, pay-for-performance schemes are increasingly being used to incentivise the quality of clinical processes and best practice in the NHS. International experience, together with evidence from the Quality and Outcomes Framework, advancing quality, and best practice tariffs, supports a role for the payment system in improving quality. But, NHS pay-for-performance schemes have not been universally successful; in particular the evaluation of Commissioning for Quality and Innovation (CQUIN) schemes has been more negative.

Features associated with the success of pay-for-performance schemes in the English NHS have been:

- a clear evidence base
- clinical engagement and support
- sufficient longevity to encourage investment in change by providers
- feasibility in practice
- simplicity.

The most successful schemes have also included non-payment quality improvement measures, such as shared learning and public reporting of data.

Evidence from evaluations of pay-for-performance schemes in English hospitals also suggests that while the definition of goals and targets can often be done best at a local level, and may help clinical engagement, developing the technical aspects of payment and pricing requires specific expertise and use of evidence, for which local areas may not have the capacity or expertise.

Although the ultimate purpose of the health care system is to improve patient outcomes, there is currently limited evidence to support financial incentivisation of outcomes. There has, however, been only limited experimentation and even scarcer robust evaluation, in part due to the fact that outcomes are far more difficult to measure and attribute than processes of care. For something to be incentivised, it must be both measurable and directly attributed to the provider. Outcomes are often difficult to measure, distant in time from the care activity, and influenced by multiple determinants, including many outside the control of the health sector, making attribution to specific provider actions problematic. There are also inherent risks to incentivising outcomes, including to equity and equality of access to care.

Best practice care for many conditions, including long-term chronic conditions, requires coordinated action across multiple health and non-health organisations and individuals. In its current form, the payment system does not support joint working between organisations within the health service, let alone more widely. Both providers and commissioners of health care are fragmented, with separate budgets and payment systems for different services, which act as a barrier to joint working and integration of treatment pathways (Shaw and others, 2011). Evaluations of the current experimentation with pathway payments and 'year of care' approaches will help build the evidence as to how payment approaches might best support integration, and findings are starting to emerge from some international pilots of this approach.

Predictability and stability are also important to commissioners and providers, and to the sustainability of local health economies. Volatility and instability in prices and incentives impedes long-term planning and investment in services, and reduces compliance with the payment system, as has been seen with CQUIN and Payment by Results. While some flexibility and movement will be essential, signalling this over a longer time period will be of help to both commissioners and providers.

Different services will need different payment methods. For example, while it may be appropriate to incentivise a process where it is directly linked to an outcome, more complex outcomes with multiple determinants will need a different approach. In addition to financial incentives, thought should be given as to whether objectives are better achieved through other levers.

In this paper, we describe the current payment systems in operation for NHS-funded care, and review the evidence from their evaluation. In our accompanying paper, *NHS payment reform: lessons from the past and directions for the future*, we make recommendations to NHS England and the sector regulator, Monitor, in their work to reform the NHS payment system.

1. Introduction

The National Health Service (NHS) faces an unprecedented challenge in meeting growing demand for, and costs of, health care in the context of tightly constrained resources (Roberts and others, 2012). With a broadly static health budget in real terms over the five years to 2015/16, the NHS in England is aiming to make efficiency savings of four per cent a year over this period, equating to a total of £15 to £20 billion. Sources of these rising financial pressures are a growing and ageing population, increasing prevalence of chronic conditions, and rising costs of health care. With further fiscal consolidation planned to at least 2017/18 (HM Treasury, 2013), it looks increasingly likely that the current period of austerity will extend for the remainder of the decade. The Nuffield Trust has previously estimated that an ongoing freeze of the health budget over the decade would require continued efficiency savings of four per cent a year (Roberts and others, 2012).

One tool that can be used to promote efficiency is the payment system through which NHS commissioning bodies purchase health care from hospitals, GPs and other providers. National policy-makers and local commissioners are exploring how changes in the structure of payment systems for health care can be used to improve the efficiency of the NHS. Efficiency is measured by comparing the outputs of the health service with the inputs used, or the cost of providing care. Conceptually, the measure of outputs should include the quality of care, as well as the volume of care. However, the measurement of outputs in health care is complex and there are concerns that quality differences are not effectively captured. Measures of efficiency of health services are therefore often a simple comparison of activity and cost, rather than quality-adjusted output. However, while the NHS seeks to improve efficiency, it must also maintain the availability of care to the population, and achieve its broad ambitions, as set out in the Department of Health's 2012 Mandate to NHS England. These include providing safe and high-quality care, reducing premature deaths, helping people with long-term conditions to remain independent, 'joining up' care across providers, and improving experience of care. It is intended that the best health outcomes will be achieved through strengthening the autonomy of local commissioners and providers, and allowing them to innovate (Department of Health, 2012d).

NHS England's budget is for the purchase of health care services in line with these ambitions for the NHS. The system through which this health care is paid is one of the central policy tools available to NHS England to influence hospital and other health care providers' behaviour.

While the payment system is at the centre of attempts to improve the crude efficiency of health care, it must also promote, or at the very least support, these wider system objectives. It is generally recognised that reform to the payment system for NHS care is needed to support the creation of a health care system that is sustainable and aligned with the needs of today's population. Indeed, an additional objective set for NHS England by the Department of Health is to have made 'significant improvements in the payment system by 2015, so that it is transparent, and rewards providers for "doing the right thing"' (Department of Health, 2012d). However, with numerous approaches available

to pay for health care – each of which carries advantages, disadvantages and potential adverse effects – this is a highly complex task.

The 2012 Health and Social Care Act transferred responsibility for the design of the NHS payment system from the Department of Health to NHS England and the sector regulator, Monitor (NHS Confederation, undated). NHS England specifies the services to be priced, determines the design of currencies, and sets rules about local variations to the national price where services are uneconomic. Monitor is responsible for developing the methods for setting prices and for calculating prices for the services included in the national tariff (price list), as well as setting rules for local pricing. This report reviews some of the recent evidence on provider payment reform, and aims to draw lessons from research in the UK and other countries for the scope and direction of a medium-term strategy for reforming the provider payment systems in the English NHS. We examine the role of the payment system in achieving the overarching objectives of the NHS, and we assess what payment approaches can achieve, what the limitations are, and what reform to the system must address.

Approaches to paying for health care

The system used to pay for health care encompasses both the unit of service for which payment is made (the currency) and the price paid for that service. Many factors can be varied in the design of a payment system for health care, and the optimal design will depend on the objectives of the health care system. For instance, payments can be made to the provider to:

- cover a range of services for a specified time period (block budget)
- care for a specific patient or population (capitation)
- provide specific services (fee-for-service)
- provide a specified quality of processes or outcomes of care (performance-related pay).

Payments to providers may be made prospectively as a fixed amount, based on assessment of local patient needs and prediction of services needed by the population served. In these cases, agreements will be needed between commissioners and providers as to where the risk and benefit fall if actual provision is above or below the predicted level. Alternatively, payment may be retrospective, reimbursing providers for actual services provided following the event; in this case, there may be an upper limit on the amount of service that will be paid for. Payments for a service may be directly proportional to the units of that service or may be conditional on reaching a threshold or target level, and the unit payment may change based on volume. The level of payment may be fixed or subject to negotiation between commissioners and providers, and may also vary depending on the characteristics of the provider or the patient seen. Payments may be withheld or reduced for non-compliance.

Varying any of these parameters can provide incentives, or indeed disincentives, to influence the behaviour of providers. In turn, changes in provider behaviour affect the costs incurred by them and the quality of care (Scott and others, 2011). Payment systems are therefore used in attempts to influence the achievement of objectives such as quality, efficiency and cost control. A range of payment approaches are used by commissioners of health care, both within the English NHS and across international systems (see Box 1.1).

Box 1.1: Payment methods in health care

The main payment methods in health care can be ordered by the extent to which they 'bundle' together payments for services. At one end of this spectrum are block budgets; at the other, fee-for-service payments.

Block budgets

The payment for all services to be provided is bundled together, and a prospective lump sum is paid to a provider at defined intervals, independent of the number of patients treated or the amount of activity undertaken. A block budget provides an overall spending limit that will constrain the volume and/or quality of the services provided (Dredge, 2004). Under this arrangement, the provider bears the risk for increased demand and cost of care, and the commissioner for decreases, unless there are arrangements to share risk or surplus. The ability of a block budget to achieve policy objectives will be dependent on contractual conditions of the payment (for example, around quality, efficiency and volume) and also on the proportion of the provider's total revenue included in the block. Salary payments, in which the periodic lump sum for bundled services provided is paid to an individual, are similar to block budgets.

Advantages:

- Transaction costs are low.
- Expenditure/income is predictable if the budget has a fixed cap and no further payment can be made for additional costs incurred by the provider.
- Provides flexibility for provider innovation – where this is cost-neutral or cost-reducing – as providers can change the service without it having a direct impact on their income.

Disadvantages:

- Lacks transparency and accountability.
- Increases in activity are disincentivised.
- Cost-increasing breakthrough innovation is disincentivised; added to this is the constrained access that providers have to capital finance, and flexibility to innovate where up-front investment is required, is limited.
- Excess demand may cause providers to ration services or result in a decline in the quality of care, without additional mechanisms to sustain volume or quality, or improve efficiency.
- Providers may avoid or under-serve costly, high-need and complex patients.
- Choice and competition are not supported as money does not follow the patient.
- 'Better' providers will attract more work, but not more resources; conversely, providers performing less well may attract fewer patients, but would not lose resources.

Capitation

Prospective, periodic, lump-sum payments are made to a provider or a network of providers per enrolled patient, for a range of bundled, specified services. Ideally, capitated budgets are 'weighted' (risk-adjusted) to take account of the fact that some patients require additional, or more costly, services.

Advantages:

- Transaction costs are low, although weighting adds cost.
- Cost containment and financial control are supported.
- Providers are incentivised to attract more patients as money follows the patient, which may in principle incentivise improved quality in dimensions of care that patients value and can observe.

Box 1.1: Payment methods in health care (*continued*)

Disadvantages:

- There is no incentive to provide additional or more costly services for patients enrolled.
- If there is no patient choice of provider, capitation funding provides no incentive for providers to be responsive to patients. In fact, it can create an incentive for providers to discourage patient utilisation by being unresponsive.
- If payments are not fully risk-adjusted, providers may avoid patients with high levels of need, or those whose needs are under-compensated for by the weighting formula.
- If the capitation payment covers only part of the patient's health care (for example, primary care), there is an incentive for providers to shift more care to other services and limit the range of care they provide.
- While in principle providers have some incentive to invest in prevention, this is rarely the case in practice. Single-year contracts are not sufficiently long for potential savings from preventive interventions to be realised by providers; hence investment in these is not encouraged, and providers have stronger incentives to save costs by lowering quantity or quality of services. Contracts either do not adequately specify or enforce minimum standards, and additional payment mechanisms are increasingly being used alongside capitated contracts to incentivise high-priority preventive interventions.

Case-based payments

Providers are paid a fixed sum for an episode of care, based on groupings of clinically similar diagnoses or procedures that entail similar costs. This method involves less bundling than capitation payments, as reimbursement is for an episode of care, rather than a period which may or may not include activity. Some bundling remains, however, as an episode may include multiple activities. Similarly, payments may be made for a defined pathway of care, for a patient with a particular diagnosis. Risk is apportioned between commissioners and providers: in principle, 'epidemiological' risk (arising from variations in the incidence of disease) falls on the commissioner; while 'clinical' risk (associated with what is done to the patient) falls on the provider. However, this distinction can break down in the presence of supplier-induced demand or billable readmissions caused by avoidable errors in care or complications.

Advantages:

- Patient choice and competition are supported as money follows episodes of care.
- Improvements to quality may be incentivised through patient choice.
- Grouping episodes by diagnosis can facilitate comparisons of clinical quality to inform choice and also facilitates benchmarking of costs.
- Providers are incentivised to reduce cost per episode (an advantage only if this is achieved through productivity improvements rather than a decline in quality).
- Increases in activity are incentivised (only an advantage if this is cost-effective and appropriate activity, for instance to reduce waiting times) (Geissler and others, 2011).

Disadvantages:

- The incentive to treat more patients might stimulate unwanted as well as desired activity through 'supplier-induced demand'.
- Financial control for commissioners is more difficult, unless a limit on volume is specified.
- Transaction costs are higher due to the need for billing and more sophisticated costing systems.
- Quality may fall as a result of attempts to increase profit by reducing costs of care where prices are fixed.

Box 1.1: Payment methods in health care (*continued*)

- Providers may be disincentivised to introduce quality-raising but cost-increasing new technologies (Quentin and others, 2011).
- Providers may select the least complicated patients, who are likely to cost less to treat.
- Providers may 'up-code' classification of patients into a more highly reimbursed group.
- Service innovation may be more difficult as currency and payment levels reflect past models of care and costs.

Fee-for-service

Payment is made retrospectively to providers for each unit of service provided, in other words, each activity or patient contact, according to a fixed price schedule. All the risk for increasing cost falls on the commissioner.

Advantages:

- Supports patient choice and competition, and thereby possibly increases in quality.
- Could promote equity of care, as providers are paid for all treatment they choose to provide.
- Supports quality and comprehensive care as the provider has no incentive to withhold or skimp on care.
- Rapidly supports innovation that expands or changes the use of treatments and technologies already on the reimbursement list, which can be reimbursed quickly.

Disadvantages:

- Highly challenging to financial control and likely to increase spending through increases in activity as a result of supplier-led demand, or through their specific treatment decisions.
- Providers are not incentivised to improve efficiency or to work jointly with other providers.
- Providers are not incentivised to prevent future ill health, unless preventive interventions are specifically paid for, with a sufficiently attractive margin. Moreover, there is no incentive to take a population-level approach to prevention.
- Can delay innovations that require addition of a new technology to the reimbursement list, as control of entry to the list is a mechanism for control of expenditure under fee-for-service.

Design of a payment system for health care

All of the payment methods described in Box 1.1 have their advantages and disadvantages, and the choice of method will depend on the context and the objectives to be achieved. No single method is capable of suiting all purposes, hence it is common to mix or 'blend' different methods within a payment system, so that some payment is prospective and some is cost-based or retrospective (Ellis and McGuire, 1990; 1993). The prospective element can be used to incentivise providers to exercise appropriate economy in the supply of care, while retaining a retrospective element can enable commissioners to incentivise specified interventions and mitigate against patient selection, which can arise when epidemiological risk falls on the provider. No single payment method is effective in maximising cost-effectiveness at an overall system level, and this requires added complexity in a payment system, with complementary management and contracting levers.

What is also apparent is that any of the methods described could have unintended adverse consequences. Anticipation of these, plus additional measures to ensure that the signals are clear, can help to ensure the payment mechanism has its intended effect. For example, activity-based payments are frequently blended with other methods and controls to mitigate against adverse effects:

- provision can be made for additional payment for exceptionally high-cost patients to protect equity of access to care
- coding can be audited to detect up-coding
- the incentive to provide unwanted activity can be countered by paying lower (or even zero) prices for activity above a specified level.

To encourage innovation, extra payments can be made over the short term for technologies that increase costs, ideally after establishing cost-effectiveness. There are particular risks to quality of care, especially where cost savings are being sought. It is therefore common to introduce pay-for-performance elements into payment systems, which reward or penalise providers for aspects of their performance. They can be used in conjunction with all of the payment methods described above, through imposing standards or criteria contractually, and making payments align with these.

For example, to incentivise quality, providers may be rewarded for compliance with evidence-based guidelines or standards by being paid a higher price for ‘best practice’ or better patient outcomes; alternatively, part of their payment may be conditional on achievement of a certain level of outcomes, or they may be paid less (or not at all) for poor outcomes or adverse events. For cost containment, payment for activity above a specified level may be at a fraction of the standard rate; also, providers may be rewarded for productivity improvements by being allowed to re-deploy or profit from savings generated, conditional upon maintenance of quality (Mannion and Davies, 2008). There are, however, a number of difficulties with implementing pay-for-performance, and careful considerations must be made to avoid unintended effects (see Box 1.2).

Objectives for a reformed payment system for NHS care in England

Monitor and NHS England’s objectives for the redesigned payment system are that it will:

- reimburse delivery of outcomes rather than processes
- support changing patterns of care
- ensure efficient allocation of resources
- account for links between health and other sectors (Monitor and NHS England, 2013a).

These are ambitious objectives. While the organisations acknowledge the potential for adverse consequences of financial incentives, as well as the need for clear signals that are consistent with other forms of incentive (Monitor and NHS England, 2013a), are they at risk of over-loading the payment system through overly ambitious objectives?

While payment systems are important levers for influencing provider behaviour, as discussed above, there are disadvantages and risks associated with all approaches, some of which may act as a barrier to achievement of health care objectives. Balancing and mitigating these to achieve the desired signals is highly complex. To add to the complexity, the payment system is just one among multiple incentive tools and other factors that influence provider behaviour. These can act synergistically with the payment

Box 1.2: Key considerations when designing pay-for-performance schemes

- The definition of 'performance' and whether to link payments to absolute or relative levels. Plus, if the latter, whether this is comparative between providers, or related to change within a provider over time.
- The ability to measure the aspects of performance that are of importance. Delivery of health care requires multiple actions, some of which can be observed and measured and others not/less well. Rewarding performance first requires the ability to measure it.
- Selection of performance targets in areas with known room for improvement, as the impact and hence the value for money may be insignificant in areas where extensive quality improvement work has recently been undertaken (Sutton and others, 2012).
- The problem of attention shift (Holmstrom and Milgrom, 1991). There is the risk that unmeasured/unrewarded work may be sacrificed for measured/rewarded work, which could have negative consequences for overall efficiency and patient outcomes. This is particularly concerning where pay-for-performance targets are determined by what can be measured, rather than what is of direct importance.
- Pay-for-performance frequently rewards compliance with processes of care, rather than outcomes. This is largely because processes are easier to measure and outcomes often distant in time and difficult to attribute to a single activity. In this way, pay-for-performance can become equivalent to fee-for-service.
- If rewards are based on outcomes, however, and only partial adjustments can be made for patient characteristics, providers may avoid treating sicker patients or those with complex conditions.
- Extrinsic motivation provided by incentive schemes can reduce intrinsic motivation, which is generally high among health professionals.
- Pay-for-performance may act as a barrier to creativity and innovation by increasing the financial risk associated with these, and thereby can encourage practice by rote.

system, or may contradict and negate signals provided through payment. It is therefore important that different priorities and levers are aligned.

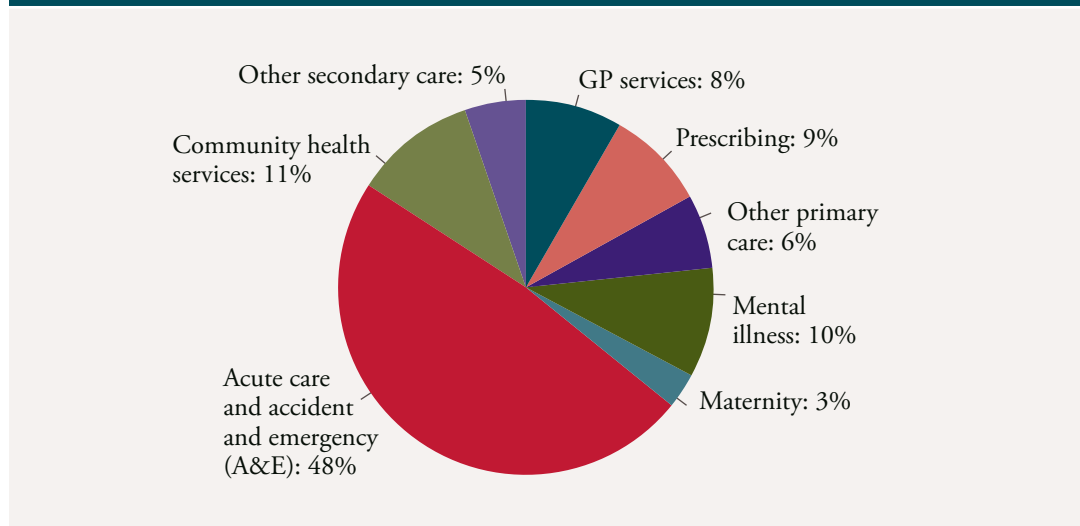
The optimal level of regulation of the payment system is also to be determined. This can range from supporting local payment decisions by improving the quality of information available or provision of guidance, through setting rules to constrain local negotiations, to enforcing national prices for centrally specified units of purchase (Monitor and NHS England, 2013a).

2. Payment systems in the English NHS

Who pays for what?

The NHS is funded through general taxation, with limited direct payments by individual patients, primarily for pharmaceuticals, dental and ophthalmic services. Until March 2013, the majority of funding for health services in England (over 80 per cent) was allocated from the Department of Health to local area primary care trusts (PCTs), who were responsible for commissioning health care and public health services for their populations. Figure 2.1 shows the distribution of PCT expenditure across primary and secondary care services in 2012/13. In this final year that PCTs existed, their total allocation was £91.6 billion. By far the largest spend (48 per cent of the total) is on acute and emergency services.

Figure 2.1: Health care purchased by PCTs in 2012/13



Source: Department of Health, 2013a.

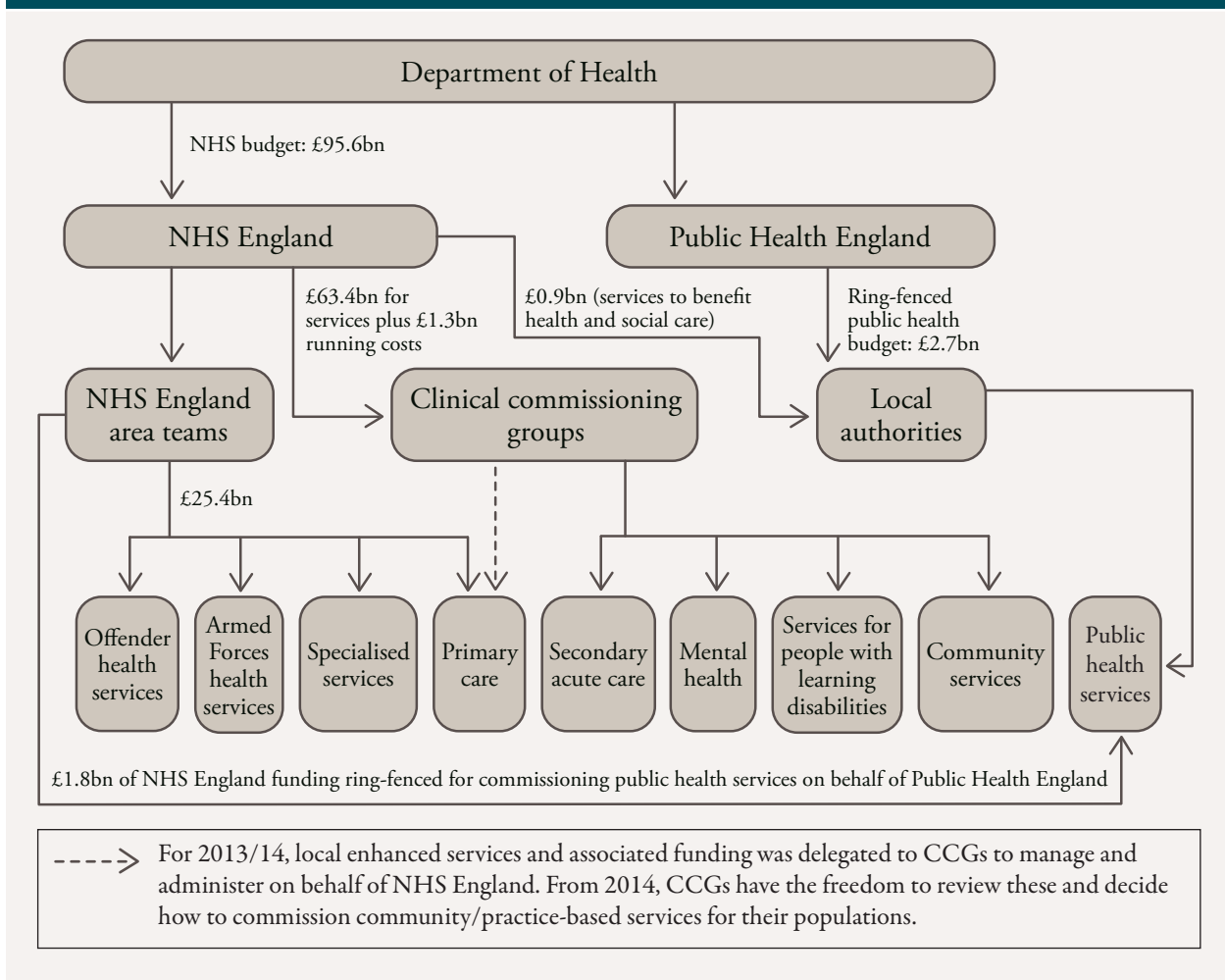
The reforms brought about by the 2012 Health and Social Care Act dramatically changed the commissioning arrangements for health care in England, as illustrated in Figure 2.2. Funding for health care services is now transferred from the Department of Health to NHS England, an independent arm's length body (NHS England, 2013a). NHS England's revenue budget for the purchase of NHS services in 2013/14 is £95.6 billion, including £1.8 billion for specific public health functions. NHS England is organised into four regional centres, beneath which operate 27 area teams. These area teams have responsibility for directly commissioning £25.4 billion of health care services, including general practitioner (GP) and other independent primary care contracts (including dentists and ophthalmologists) for their local populations (£11.1 billion), and specialised services (£12.0 billion; high-cost/low-volume services, led by ten of the area teams for national consistency) (NHS England, 2013b).

NHS England allocates resources to the 211 clinical commissioning groups (CCGs). These newly formed clinically led organisations replaced PCTs as commissioners of

non-specialist health care services for their local populations on 1 April 2013. CCGs now control around two thirds of the NHS budget in England (£64.7 billion), with which they meet the costs of non-hospital prescription drugs and commission non-specialist secondary care and community services for their local populations (Naylor and others, 2013).

The third statutory group of commissioners of services in the reformed health system in England are local authorities. Directors of Public Health and their teams are now based in local government, and local authorities are now responsible for commissioning public health services for local populations, holding a £2.7 billion public health budget (separate from the public health funding held by NHS England), in addition to their budgets for social care, housing, education and other local services (Naylor and others, 2013).

Figure 2.2: The commissioning landscape in England since 1 April 2013 (with 2013/14 budgets where shown)



It is apparent, therefore, that the health system reforms have fragmented commissioning arrangements in England, with several separate organisations now responsible for commissioning the services that were previously commissioned by PCTs. Commissioning of primary care services is now performed by a separate organisation from acute, community and mental health services, and the majority of public health services are commissioned by further separate, non-NHS organisations. This poses a significant

challenge to ambitions to integrate services better across sectors. Not only do providers remain separate, but now there is added fragmentation of commissioners, and hence budgets, for these services. The challenge now is to take account of this fragmentation in designing a payment system for NHS care which supports joint working and the creation of a true health care system for patients.

Evolution of payment systems in the English NHS

Over the last decade, the payment system for NHS care in England has evolved away from one in which funding was predominantly transferred from commissioners to providers through block grants and contracts with locally determined prices (Monitor and NHS England, 2013a). These arrangements gave little incentive to improve the quality or efficiency of care, and a series of changes to the payment system has since sought to focus providers on improving quality, efficiency and waiting times.

Primary care services

GP services will account for around £6.4 billion of NHS funding in England in 2013/14 (NHS England, 2013b). Historically, most GPs were self-employed, and paid individually under a General Medical Services (GMS) contract, with payments based on:

- a mix of weighted capitation, to take account of relative levels of need of practice populations
- practice allowance, under which expenses were reimbursed at the national average level, giving incentive to economise
- elements of fee-for-service for certain specific services such as immunisations, minor surgery and out-of-hours visits.

The late 1990s saw a crisis in GP recruitment and retention, and growing concerns about the quality and equity of primary care (Doran and Rowland, 2010), triggering reform to the GP contract.

The Personal Medical Services (PMS) contract, introduced in 1998, was held with practices rather than individual GPs. The contract allowed local variations to the standard GMS contract to be negotiated, enabling greater targeting of services to local needs and the setting of quality standards. PMS practices are paid a monthly lump sum, based on an equivalent GMS contract, with expenses met out of this income. This arrangement supported employment of salaried doctors and other practice staff, with the contracting GP partners being allocated an agreed share of the profits, as payments resembling a monthly salary. By 2006/07, about 40 per cent of practices in England were contracted under PMS rather than GMS arrangements (National Audit Office, 2008), and by 2010, about one fifth of GPs had chosen salaried status.

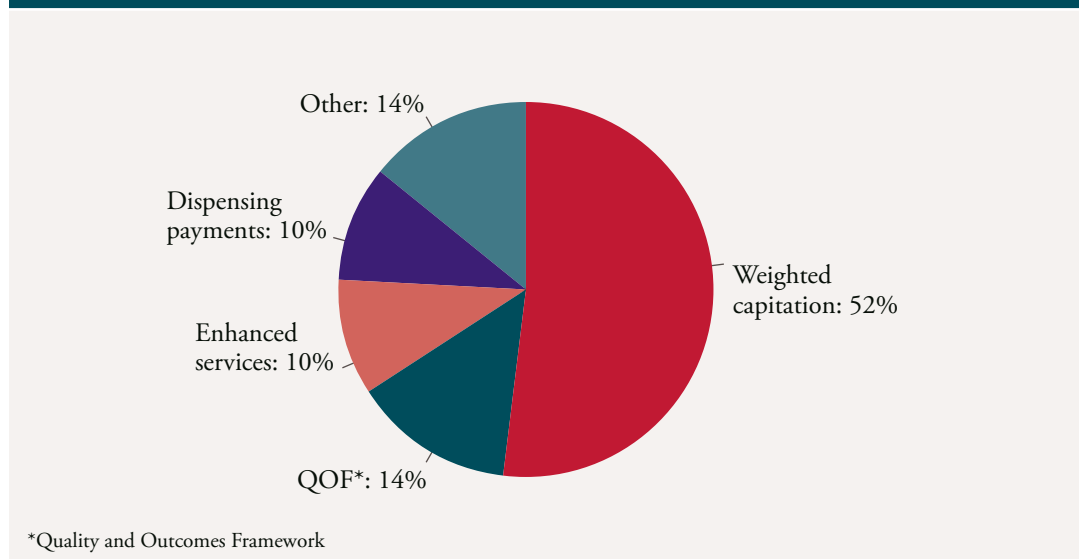
The GP payment system changed in 2004, with major reforms to the GMS contract. The new contract was based on practices rather than individual GPs, and was funded out of a fixed national global sum for primary medical care. It incorporated a voluntary pay-for-performance component, the Quality and Outcomes Framework (QOF), and more refined weighted capitation payments. It also provided the opportunity for commissioning 'enhanced' services from practices; these being additional, specialised or innovative services aimed, for example, at meeting specific local needs, supporting patient choice of hospital, or reducing demand on secondary care. As with PMS contracts, practice expenses were paid out of gross income. Any practice threatened with loss of income from core services under the new contract had its income brought up to the historic level under a 'minimum practice income guarantee', which inhibited

reallocation of primary care funds to under-doctored or under-funded areas. At this time, PCTs were also given powers to commission primary medical services from alternative providers, such as private companies, especially in under-served areas.

The aims of the QOF were to improve quality of primary care, embed preventive measures and stimulate improvement in chronic disease management to reduce avoidable hospital admissions (Health Foundation, 2011). Up to 25 per cent of practice income was linked to achievement of quality targets for several chronic conditions (Doran and others, 2011). About half of these related to adherence to clinical processes, and the majority of the remainder to measuring patient experience and achieving desired organisational practices, such as recording and reporting of activity and quality. Only a small proportion of payment was conditional on achieving clinical outcomes. A maximum of 1,000 points are available, with an average payment of £130 per point in 2011/12. Most practices reached most targets rapidly, exceeding Department of Health expectations and, as a result, the QOF rapidly raised incomes. PMS practices are also able to participate in QOF, either using the national framework or negotiating local arrangements (Health & Social Care Information Centre, 2013).

Figure 2.3 shows the mix of payment systems used for GP services – weighted capitation is the largest element but the system includes a significant pay-for-performance component (QOF) and some fee-for-service (Pike, 2010).

Figure 2.3: Funding GP practices in England



Note regarding dispensing payments: Not all GPs in England are permitted to dispense, with only certain doctors being permitted to provide pharmaceutical services to patients meeting certain circumstances. These circumstances are that the patient would have serious difficulty in obtaining the necessary drugs or appliances from an NHS pharmacist due to distance or inadequate means of communication, or if the patient lives in a rural area more than a mile away from a pharmacy.¹ Only around 15 per cent of practices dispense, and the proportion of a practice's income derived from dispensing services varies greatly, but may be as high as 50 per cent in the most rural and remote practices, and in some practices dispensing income may cross-subsidise medical services.²

1. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/212872/Chapter-15-dispensing-doctors-services.pdf

2. www.nhshistory.net/gppay.pdf

For the 2013/14 GP contract, QOF thresholds have been raised to further improve performance and new indicators have been added. The National Institute for Health and Care Excellence (NICE) has a new role in recommending QOF indicators, producing a menu of evidence-based, clinically and cost-effective indicators – cost-effectiveness has not previously been taken into account in QOF. Clinical areas are being prioritised by NICE, with suitability considered according to:

- prevalence
- accuracy of data extraction
- clarity of diagnosis
- relevance of incentivised actions
- how directly change can be attributed to primary care
- possible unintended consequences.

These criteria do, however, mean that straightforward technical and pharmaceutical interventions are prioritised over more complex community-based interventions (Gillam, 2013). In further contract changes, the minimum practice income guarantee is being phased out, and new directly enhanced services introduced, including care management for frail older people, early diagnosis of dementia, and telehealth and online access.

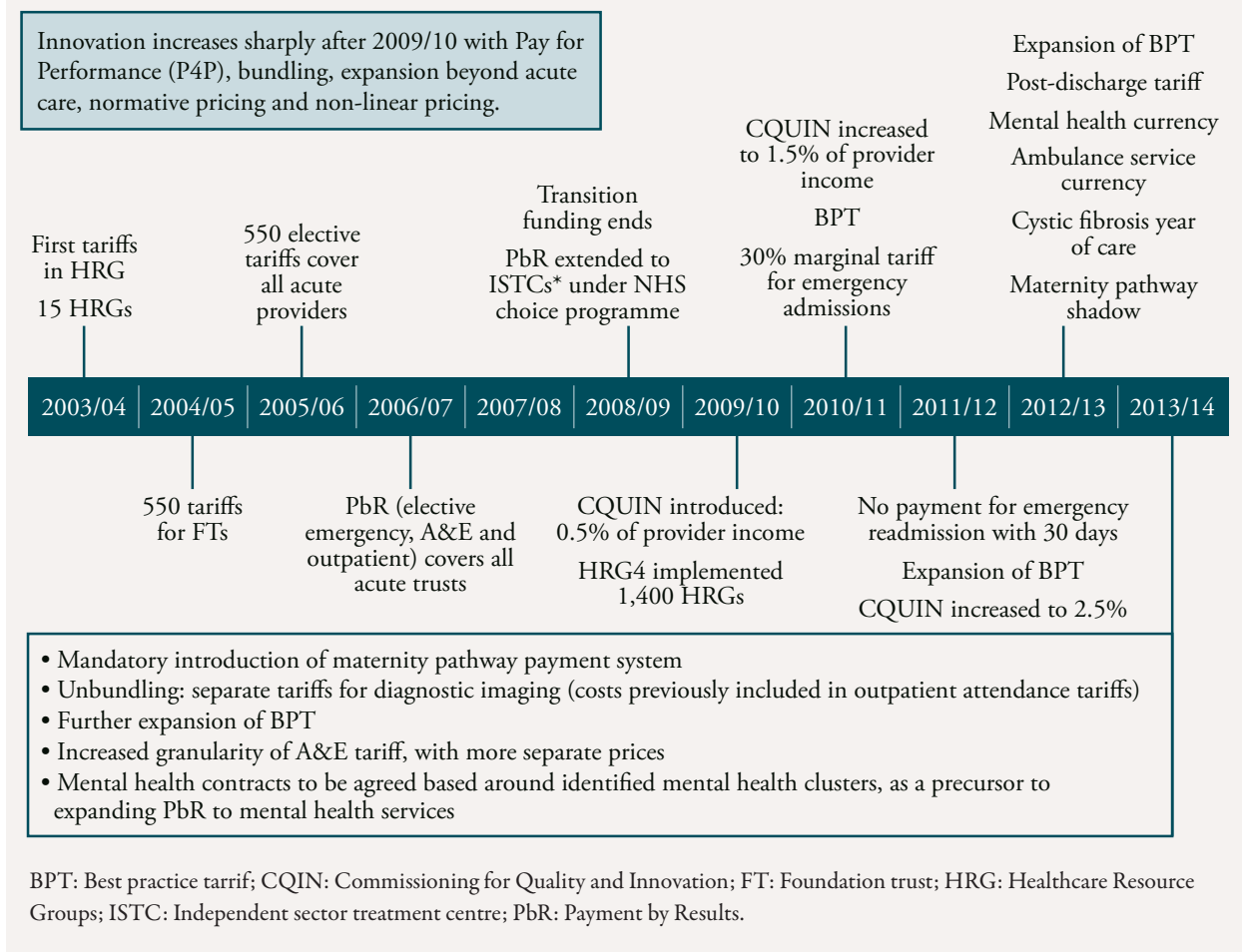
Secondary care: acute services

PCT allocations for acute services were based on a weighted capitation formula to identify the area's 'fair share' of the overall NHS budget, based on relative need and the policy on how quickly areas should move from their historic funding level to the fair share (the pace of change policy). NHS England is currently undertaking a review of resource allocation for CCGs and centrally commissioned services (NHS England, 2013b). While the predominant determinant of CCG allocation is likely to be the population they serve (capitation), NHS England will also use a 'quality premium' to reward CCGs for improvements in the quality of services commissioned and associated improvements in health outcomes, and reductions in inequalities. These performance-related payments will be contingent on four national measures (reducing potential years of life lost through health care amenable mortality, reducing avoidable admissions, improving patient experience of hospital services, and preventing health care associated infections) plus three local measures based around local priorities and agreed between CCGs and NHS England area teams (NHS England, 2012). By far the biggest portion of the CCGs' budget is spent on acute and emergency secondary care services.

There have been major reforms to payment for acute and emergency services over the last decade. Before this time, hospitals were paid predominantly under block contracts for a broad range of services (Farrar and others, 2007). These commonly specified minimum and maximum levels of provision, with activity above or below these levels triggering actions such as renegotiation or data validation, and the volume of care provided in a year would influence year-to-year contract renegotiation where funding growth permitted (Farrar and others, 2007). There was little incentive for providers to be responsive to patients or to improve the quality or efficiency of care. In 2002, the government committed to large and prolonged investment in the NHS, and reintroduced market-like purchasing of hospital care with the aim of improving patient choice. A series of aligned reforms was carried out over the following years (Department of Health, 2005), including the introduction of Payment by Results in 2003/04. This initially financed a small proportion of inpatient elective hospital care in 2003/04; was expanded to cover all

elective care by 2005/06 (Farrar and others, 2007); and by 2006/07, covered most acute activity, including non-elective, outpatient and A&E care (Department of Health, 2012a). Payment by Results remains the dominant payment system for hospital care, with £29 billion of activity covered by the tariff in 2011/12, representing 40 per cent of spending on secondary care, and covering around 60 per cent of an average hospital's activity (Department of Health, 2012a). Figure 2.4 shows the key points in the development of the Payment by Results system.

Figure 2.4: The evolution of Payment by Results in the NHS



Payment by Results uses a nationally fixed diagnosis-related case-based tariff to reimburse hospitals for the amount and type of care provided (Farrar and others, 2007), making a link between both the volume and case-mix of hospital activity and income. The aims were to effect changes in the efficiency and quality of care in English NHS hospitals, and to increase activity at a time of long and growing waiting times for elective care (Department of Health, 2012b). Payment by Results meant money would 'follow' the patient and, because prices were fixed, competition for patients would be on the basis of quality rather than price. Its introduction was accompanied by commissioning, which supported and supplemented patient choice of provider, with private providers being permitted to compete for elective surgery.

Under Payment by Results, providers are reimbursed for 'spells' of activity. Spells, covering the period from admission to discharge, are coded as Healthcare Resource

Groups (HRG), based on types of patient and treatments with similar cost implications. There are currently more than 1,300 HRGs included in the national tariff. In the English system, a single hospital stay, or spell, may include multiple consultant episodes. Where this is the case, the dominant episode counted within the spell will determine the HRG code for billing of that spell. Tariff prices are based on national average provider costs (as submitted by providers), which are adjusted equally for all providers to reflect changes in costs over time (for example, due to inflation, technology and efficiency improvements). Some tariffs are also adjusted to take account of NICE guidelines on cost-effective technology. Finally, tariffs are adjusted using the market forces factor to give a local price for a trust that reflects unavoidable local differences in costs (Farrar and others, 2007).

With the end of the period of rapid financial expansion of the NHS in 2010/11, the potential of Payment by Results to generate efficiency savings in hospitals has been explored. The annual uplift in prices was set at zero per cent in 2010/11, at a time when the inflationary impact of hospital pay and prices was expected to run at about 3.5 per cent a year. In 2010/11, a marginal rate was applied to emergency admissions above the 2008/09 volume, for which only 30 per cent of the tariff price was paid, and seven 'never-events' for which hospitals would receive no payment listed (such as wrong-site surgery) (Department of Health, 2009). In a further change to reduce costs and incentivise quality in emergency care and follow-up, responsibility and funding for patients in the 30 days following discharge was passed to acute providers in April 2012, with no payment for emergency readmissions over this period (Department of Health, 2011c).

Since 2011, the prices paid under Payment by Results have fallen, as part of the Department of Health's strategy to deliver efficiency savings of £15 to £20 billion by 2015. Table 2.1 shows the annual calculation for the tariff uplift between 2011/12 and 2014/15.

Table 2.1: The components of the annual change in Payment by Result prices

	2011/12	2012/13	2013/14	2014/15
Pay and prices	2.4%	2.0%	2.5%	1.9%
Revenue cost of capital	0.2%	0.2%	0.2%	0.2%
Service development	0.0%	0.0%	0.1%	0.4%
Clinical Negligence Scheme for Trusts (CNST)	-	-	-	0.1%
Efficiency factor	-4.0%	-4.0%	-4.0%	-4.0%
Net price adjustment	-1.5%	-1.8%	-1.3%	-1.5%

Source: Monitor and NHS England, 2013b; Department of Health, 2010; 2012a; 2013b.

A concern with activity-based payment such as Payment by Results is that quality may suffer if providers cut costs to remain at or below the price paid. This becomes particularly pertinent with reductions to tariff. To guard against this, attempts have been made to incorporate pay-for-performance elements alongside Payment by Results (see Box 2.1).

Box 2.1: Pay-for-performance schemes in NHS hospitals

Advancing Quality

In 2008, this scheme was introduced in 24 NHS hospitals in the north west of England. It was a tournament-style scheme, based on the US Hospital Quality Incentive Demonstration. Hospitals submitted data on 28 quality measures, including mortality and readmission rates, in five clinical areas (Sutton and others, 2012). At the end of Year 1, quality scores in the top quartile earned a bonus payment of four per cent of Payment by Results revenue for that activity, with hospitals in the second quartile receiving two per cent. Quality improvement was supported by non-financial mechanisms, including feedback on performance, support to standardise data, shared learning events, public reporting of results, plus additional internal activities. Participating hospitals all agreed to allocate bonuses – to the clinical teams whose performance had earned them – for investment in improved clinical care.

In Year 2, hospitals earned bonuses for:

- ‘attainment’ if quality exceeded the median level in Year 1
- ‘improvement’ if the increase in quality was in the top quartile
- ‘achievement’ if quality was in the top or second quartile.

The scheme was halted halfway into Year 2, and replaced by the national pay-for-performance scheme, which involved withholding payments, rather than earning bonuses.

Commissioning for Quality and Innovation (CQUIN)

This payment framework was introduced across England in 2009, with the aim of aligning financial incentives with health system goals. CQUIN covers ambulance, community, mental health and learning disability services, alongside acute hospitals. Under CQUIN, a proportion of provider income is conditional on achievement of locally agreed quality and innovation goals (although national goals have now been added for acute providers). The CQUIN payment covers all income, not just payments under Payment by Results tariffs. The incentive was initially 0.5 per cent of provider income (actual outturn value of the contract, including tariff, non-tariff and cost-per-case income), but was increased to 1.5 per cent after the first year, and to 2.5 per cent in 2012. The objectives being locally agreeing goals were to ensure quality became part of commissioner–provider negotiations, to foster shared objectives, and to involve providers in developing schemes. It was intended that national indicators would be used to monitor performance against the local goals, although this has not been the case in practice, with a proliferation of locally developed indicators in use. NHS England reviewed the design of the CQUIN scheme for 2014/15, and published their proposals in December 2013 (NHS England, 2013c).

Best practice tariffs (BPT)

BPTs, introduced in England in April 2010, aimed to incentivise delivery of best clinical practice by adequately reimbursing high-quality care, promoting care that is both clinically and cost-effective. BPTs pay a price for episodes in accordance with specified ‘best practice’, with the aim of reducing unexplained variation in clinical quality, and increase the diffusion of best practice. A specific approach has been developed for each BPT, based on the clinical characteristics of best practice, and the availability and quality of data. The Department of Health’s criteria for BPTs are:

- high impact (that is, high volumes, significant variation in practice, or significant impact on outcomes)
- a strong evidence base on what constitutes best practice
- clinical consensus on the characteristics of best practice (Department of Health, 2012a).

Box 2.1: Pay-for-performance schemes in NHS hospitals (*continued*)

The scheme initially covered four high-volume clinical areas (McDonald and others, 2012). Of the original BPT, one was applied nationwide through the tariff, with the others subject to local agreement, as they required additional data collection. Since 2010 there has been a rapid expansion in the number of areas covered by BPTs. For example, the number of day-case procedures covered by a BPT has increased from one (gall bladder removal) in 2010/11, to 16 clinical areas in 2013/14. Similarly, BPTs for same-day emergency care first introduced in 2012/13 now cover 19 clinical scenarios (Department of Health, 2012a). BPTs cover a wide range of different services but there are some characteristics that underpin a number of the tariffs. These include: delivering care in appropriate settings (day case, outpatient, etc), reducing avoidable hospital admissions and improving quality of care.

Secondary care: mental health and community services

While almost two thirds of hospital activity is covered by activity-based payment, through the national tariff, the predominant payment system for the remaining secondary care services has been the block budget, with block contracts used to reimburse around 90 per cent of community services, and two thirds of mental health care. Local tariffs reimburse the remainder of these services (Monitor, 2012).

Mental health services accounted for ten per cent of PCT spending on services in 2012/13, and the Department of Health intends to extend Payment by Results to include mental health. A mental health currency was published in 2010/11, which proposed that contracting for mental health services should utilise bundled, period payments (care clusters) for 20 specified conditions, for durations of between four weeks and a year (Department of Health, 2011a). The clusters were created during 2012/13, and providers asked to define packages of care associated with these. Tariffs will be based on costs of these packages, as submitted by providers. Although the intention was to introduce the tariff during 2013/14, the Department of Health recognises that this is too soon, but is asking commissioners and providers to prepare (Department of Health, 2012c). In 2013/14, CCGs and providers should agree a local price for each cluster. Currencies are also being developed for a wider range of mental health services including the Improved Access to Psychological Therapies (IAPT) service and Child and Adolescent Mental Health Services (CAHMS) (Monitor and NHS England, 2013b).

The other key group of services not covered by a tariff payment system is community health services. Community health services are diverse in function and differ widely between localities across England. They include a wide range of services based outside hospitals, including care for long-term chronic conditions, preventive services, and assessment and rehabilitation services; plus community hospital services, some non-acute inpatient hospital care, and hospice care. Together, these accounted for ten per cent of PCT spending in 2010/11. Although there has been discussion about national community service tariffs, progress has been slow, hindered by local diversity of services, and lack of uniform activity and costing data (Department of Health, 2008).

Attempts have been made to extend the patient choice agenda to mental health and community services. In 2011, local areas identified at least three (of a list of eight) services in which they would implement the 'Any Qualified Provider' scheme in 2012/13 (Department of Health, 2011b). This allows patients to choose from a selection of commissioned providers when referred by their GP. Competition between 'qualified'

providers is on the basis of quality, rather than price, with all providers being paid a fixed price (NHS Supply2Health). While the national tariff sets this price for elective hospital care, commissioners will set prices locally for mental health and community services in line with guidelines from the Department of Health. Pay-for-performance aspects are added to the payment system for mental health and community services through CQUIN schemes. In 2014/15, further quality and outcomes measurement is planned for mental health services, and Monitor proposes that payments to providers could be varied according to the standards achieved for patients in the cluster. Mental health provider payments will be activity-related and some element of patient choice is being introduced.

Taking a wider view

In summary, therefore, blended payment systems are common in the NHS. The GP contract combines elements of weighted capitation, pay-for-performance and fee-for-service. For hospitals, diagnosis-related payments coexist with block budgets, retrospective per-day payments for certain outlier patients with unexpectedly long lengths of stay, and some fee-for-service for, for example, unbundled diagnostics. However, while the balance of methods within health sectors has been subject to attention and reform over the last decade, the balance across sectors is also important.

Taking a wider view of the payment system for the overall NHS budget will be an important step in developing the optimal balance of spending across services to both meet the needs of the population and maximise the cost-efficiency of provision. In doing this, it must be borne in mind that the payment approach that is most effective in achieving higher-quality care and better value is likely to vary according to the context of care.

Activity-based payment approaches such as Payment by Results work best in contexts where:

- the episode of care has a well-defined start and finish point
- the care is planned in advance
- there is evidence demonstrating the benefit and cost-effectiveness of the care to the patient.

They are helpful where priorities are enabling choice for patients, increasing productivity, and increasing utilisation of particular services, or by target population groups (where the criteria for payment must be well defined to address identified inequalities in access). As priorities differ across contexts, so will the optimal method, and the design of the NHS payment system needs to be flexible to accommodate a range of approaches that can be applied to different contexts (Monitor and NHS England, 2013a).

There has been far more development to payment for acute services than for community and mental health. Perhaps partly as a consequence of this, cost, quality and outcomes data are poorly developed for these services. In turn, this is a limitation to the development of payment systems, and must be urgently addressed to tackle imbalances across the system and barriers to policy ambitions to shift care.

Currently, the predominance of activity-based payment for acute and emergency care in theory incentivises increased activity in this sector. In parallel, the predominance of block budgets for community services and capitation for primary care services could disincentivise increased activity. Given the policy ambition to shift care out of hospital and closer to the home, this imbalance must be addressed to prevent the incentives acting

as a barrier. Neither do block budgets promote productivity improvements. If care is to be shifted into the community, productivity improvements will be needed to manage increased activity, given current budgetary limitations.

Recent local and national developments have included attempts to design payment systems that can apply across services and that support shifting of care. A number of examples involve bundling of service payments to include capitation-based 'year of care' payments for life-long conditions such as cystic fibrosis and for high-risk, multi-morbidity patients (Monitor and NHS England, 2013a), and pathway-based payments for maternity services. Another approach being explored is unbundling of payments. As part of the drive to shift care out of hospitals, the 'recovery, rehabilitation and reablement' (RRR) model separates the current acute tariff into the acute care phase and the post-acute RRR phase, allowing this to be commissioned in the community rather than hospital, where appropriate (Department of Health, 2012e). These approaches are, however, still experimental and evaluations will in future provide invaluable evidence as to the feasibility of these sorts of approaches.

In the next section, we look at evaluations of the payment systems in operation in the NHS, to assess the impact they have had in practice.

3. Review of evidence from evaluation of NHS payment programmes

With the NHS having to find savings of at least four per cent a year, the potential of the payment system to drive crude productivity savings and reduce costs is being explored. Under these conditions, it becomes more important than ever to protect against a decline in quality becoming a consequence of cuts. Financial incentives in the form of pay-for-performance schemes are increasingly being adopted in an attempt to improve the quality of care, both in the NHS and internationally. This is despite a limited evidence base to support their use.

Two Cochrane reviews published in 2011 found mixed evidence for the use of financial incentives in health care payment systems, with the evidence base being seriously limited by poor evaluation methods and reporting of the schemes. First, a review of reviews examining the effectiveness of financial incentives in changing the behaviour of health care professionals and patient outcomes included 32 studies (Flodgren and others, 2011). Overall, financial incentives were generally effective in improving processes of care, including referrals, admissions and prescribing costs. They were not generally effective, however, in improving compliance with guidelines. None of the studies examined effects of financial incentives on patient outcomes. The differences between the studies in control methods and context make it difficult to draw conclusions as to the effectiveness of different methods, however, payment for activity, specific patients or populations, or for providing a change in activity or quality of care appeared more effective than salary payments at achieving objectives.

Second, just seven studies were included in a review of evidence for the effects of financial incentives on the quality of primary health care (Scott and others, 2011). The different incentive schemes had modest and variable effects on the quality of health care provided, with greater effects found on process than outcome measures. For example, in three studies examining incentivisation of smoking cessation, referral rates and recording of smoking status were changed, but there was no impact on measures of patients' smoking cessation. Again, evaluations were poorly designed and reported, and the conclusion was that there was insufficient evidence either to support or not support use of financial incentives to improve quality of primary health care. Although systematic reviews are an important source of evidence, this methodology also has its limits. The research on payment systems examines a wide range of interventions in very complex and diverse systems.

From the available evidence then, financial incentives can influence processes of care, with little evidence of effects on patient outcomes. A major limitation, however, is the dearth of rigorously designed and reported evaluations. Moreover, the evaluation studies are focused on the short-term impact of payment systems – most examine the impact 18 months to three years after implementation. Almost none look at what happens after five or ten years, or what happens when payment systems are modified to address emerging challenges – for example, what happens when diagnosis-related group (DRG) payment is constrained within a global budget or 'cap and collar' (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) contract due to fiscal constraints?

The impact of a new payment approach will depend on the system it replaces, as well as the context into which it is introduced, including policy ambitions and other incentives and drivers at play. Factors that influence the ability of an incentive to achieve its aims include the nature and complexity of the action being incentivised, the size and method of the incentive, the health professional group and the organisational environment. These aspects are rarely examined or reported; furthermore, little data are gathered on unintended consequences, such as attention shift, gaming and loss of motivation (Glasziou and others, 2012). For these reasons, it is critical that payment schemes introduced in the NHS are subject to careful monitoring and evaluation, including of undesirable effects. Due to the very context-specific nature of the effect of payment systems, this section will focus on evidence from recent evaluations of the NHS payment systems described in the previous section. Evidence from international health care systems has been reviewed in detail in our recent paper on European payment systems (Charlesworth and others, 2012).

Primary care Quality and Outcomes Framework

Evaluation of QOF is difficult as it was implemented nationwide, leaving no obvious control group, and the multiple determinants of the chronic conditions on which QOF focuses makes attribution of changes to any one intervention difficult (Gillam, 2013). Consensus exists that QOF has improved processes in primary care and that quality of care for chronic conditions has improved since its introduction. GPs and practice nurses report improvements in teamwork, organisation and recording of care for incentivised conditions, although this has not extended to non-incentivised ones (Maisey and others, 2008). There is concern that financial incentives may have a detrimental effect on intrinsic motivation, which is typically high among health professionals. While this was not evident among GPs following the introduction of QOF, practice nurses have reported reduced motivation (Glasziou and others, 2012). Motivation was reduced, however, if health care professionals disputed the evidence base for an indicator, illustrating the importance of clinical support for incentive schemes (Maisey and others, 2008). Financial benefits of achieving QOF targets go to the practice, and hence to the GP partners, but are not transmitted to salaried staff including nurses and the growing proportion of GPs that are salaried.

The extent to which quality improvements track or exceed previous trends is unclear, with broad agreement that QOF has not had a dramatic effect overall once these trends are taken into account. Quality of primary care was improving prior to QOF. Although its introduction accelerated improvement in incentivised activities over and above trends during the first year, this came at the expense of a small detrimental effect on non-incentivised activities, for which performance was already lower, and which became relatively worse after the introduction of QOF (Doran and others, 2011; Gillam, 2013). QOF may have contributed to reducing previously wide inequalities in primary care quality. While the least deprived practices out-perform the most deprived on those QOF indicators that contribute to health gain, differences have narrowed since the introduction of QOF in 2004/05 (Dixon and others, 2010).

Incentive schemes tend to focus on aspects of care that can be measured, rather than necessarily being determined by what is of direct priority, and examples of this can be seen in QOF. For instance, prescribing and measurement feature highly among the

indicators, rather than patient outcomes, which have contributed to rising prescribing rates and costs (Gillam, 2013). Single diseases are emphasised over multi-morbid conditions, which are more complex both to measure and treat, however this does not reflect the needs and reality of the population, in which two thirds of people aged over 65 have multiple chronic conditions. There is little focus on primary prevention of ill health in QOF, necessitating commissioners to use alternative payment schemes to avoid neglect of preventive care, including locally enhanced service contracts (Dixon and others, 2010). So, while QOF incentives have led practices to organise and systematise their approach to managing chronic diseases, they have not encouraged primary prevention in general practice. There are also feelings that the focus on national targets in QOF has acted as a barrier to commissioning primary care focused on the needs of local populations, and that it does not provide appropriate incentives to practices serving populations with complex needs (Dixon and others, 2010). Some health care professionals also feel that the need to perform and record specific activities to meet QOF targets has led to consultations becoming less patient-led (Maisey and others, 2008).

There is no evidence that QOF has impacted on patient outcomes, perhaps partly as a result of difficulties in assessing this. Modelling studies have estimated the impact of changes achieved following the introduction of QOF on health, assuming that the improvements in processes directly and predictably translate to improved outcomes. A potential saving of 11 lives per 100,000 people has been estimated from the improvements in processes of care seen in the first year, with no further gain achieved in the second year due to performance typically exceeding target payment levels and plateauing by this point. Cost-effectiveness has been modelled for a few QOF indicators. Some incentive payments were found to be cost-effective, even with only modest improvements in care, although no account was taken of the costs of administering the scheme. Incremental costs per quality-adjusted life year gained ranged from £58 to £15,654; however, there was no relation between the size of payments in a clinical area and the likely resultant health gain (Gillam, 2013).

It has been demonstrated that achievement of QOF indicators is associated with some measurable reduction in costs for hospital care, possibly both within the same clinical area and others (Health Foundation, 2011). However, this study did not include the costs incurred in primary care, so could not assess the total costs of care. This is potentially an encouraging finding, as to achieve the policy ambition of better integration of care, it is important to think between care sectors with separate budgets and contracts, and to consider wider effects across the health care system as a whole.

Activity-based payment in secondary care

Evaluation of Payment by Results has found strong evidence that the introduction of the tariff payment system resulted in reductions in length of stay and increases in the proportion of day cases, across most groups of patients, providers and HRGs (Farrar and others, 2010). The attributable resource savings are estimated to be between one per cent and three per cent over a five-year period. As might be expected with a move from block budget to activity-based payment, increases in activity of between three per cent and nine per cent in the number of spells followed the introduction of Payment by Results. Increased activity was an intention of Payment by Results, to reduce waiting times for elective care, which at the time were long and growing. However, as other policies including waiting time targets were also introduced, an effect of these cannot be precluded. Evaluations of the first three and first five years of Payment by Results used hospitals in

Scotland as a control, and also made use of different timing in the introduction of Payment by Results in English hospitals (Farrar and others, 2007; 2009; 2010).

This evidence, together with the fact that there was no evidence of deterioration in the quality of care, with limited evidence of small improvements, suggests that reductions in costs were achieved as intended; through improved productivity in the delivery of care, rather than sacrifices to quality.

Other evidence suggests that Payment by Results has led to some improvements in quality through enabling patient choice, although there is no evidence to suggest Payment by Results has improved quality in the absence of this mechanism (Monitor, 2012). This evidence is consistent with the findings of studies across Europe that introduced similar DRG-based payment systems in place of block budgets. In some countries (most notably the US, but also Italy), DRG-based payment systems replaced fee-for-service methods of payment. In these countries, the change in payment system has been associated with reductions in activity (Geissler and others, 2011).

Any gains from Payment by Results need to be considered alongside any additional cost of implementing the new payment system. A study in the first regions to implement Payment by Results indicated that full implementation across England would significantly raise transaction costs to the NHS (by £40 to £60 million a year, as estimated in 2005). These additional costs were due to:

- Higher costs to providers of data collection – Payment by Results required accurate patient-level data for claiming payment; these data were previously sparse, necessitating investment in clinical coding, costing and information systems.
- Higher costs to commissioners of negotiation, monitoring and enforcement – Payment by Results provided a new incentive to hospitals to increase activity levels, with hospitals no longer having to gain approval to expand activity. While the burden of negotiating prices was removed by the national setting of prices, this incentive meant that commissioners faced greater uncertainty about what they would have to pay for, and they had to monitor volumes to determine affordability, and also the accuracy of counting and coding by providers, bringing increased potential for disagreement and dispute between commissioners and providers (Marini and Street, 2007).

The evaluation of Payment by Results in the NHS, and results from studies across Europe (see Box 3.1 on page 28), suggests that case-mix adjusted payment systems may improve the productivity and efficiency of providers of care. However, this does not necessarily mean that the health system as a whole is more efficient, and there is no clear evidence of the impact of Payment by Results on the overall efficiency of the health system (Street and others, 2011). The desirability of additional activity depends on a number of factors, including whether this activity is cost-effective (assessed as the value of the improvement in health as a result of the activity, compared with the cost). For example, productivity of provision of treatment for ambulatory care-sensitive conditions in hospitals may be achieved through incentives; however, more effective treatment and management in the community could have avoided the need for the treatment, reducing the overall cost to the NHS, and the pain and suffering for the patient.

The challenge facing the NHS in this decade is not just to improve provider efficiency, but to improve the efficiency of the health system as a whole. Despite this, the current approach to achieving the required NHS efficiency savings of £15 to £20 billion by 2015 is heavily reliant on acute hospitals improving their efficiency. In 2011/12, around half

of the efficiency savings reported by the NHS came from the acute sector (Health Select Committee, 2013). Payment by Results remains the predominant payment system for acute care, and is now being used in attempts to drive down costs and increase productivity, with tariff prices being reduced annually through the efficiency factor. Caution must be exercised to ensure that costs are not driven down too far, however, as inadequate reimbursement of costs risks a decline in quality when the volume of service is linked to payment. Indeed, the effectiveness of tariff reductions in reducing overall costs is unclear, since a rise in non-tariff revenue has been seen, compensating for the reductions.

The derivation of appropriate incentives requires good information on costs, quality and patient outcomes, which is scarce (Monitor, 2012). National tariff prices are based on providers' estimated costs for each HRG. Inaccuracies in cost data will result in reimbursement levels that do not reflect true underlying costs. This blunts financial tools, either by not incentivising efficiency, or being inadequate to cover efficient cost of provision, risking quality. A report for Monitor by PricewaterhouseCoopers (Monitor, 2012) found huge variation in costs reported, with around a third of providers reporting costs of at least 50 per cent higher or lower than the weighted average, which is used as the price paid. Some of this variation is due to disparity in how providers allocate costs to different HRGs, however, patient-level costing reveals that costs also vary widely *within* providers, driven largely by characteristics of patients, suggesting HRG codes may be grouping together patients with very different cost implications.

While accurate cost information is scarce, acute trusts are increasingly internally supplementing their costing systems with patient-level information and costing systems (PLICS), which generate estimates of costs for individual patients, based on a mainly bottom-up costing process. This enables more accurate assessment of the cost of outlying patients, and examination of clinical reasons for differences between costs and prices. Qualitative research with providers has revealed a reluctance to share these data with commissioners in contract negotiations, as it is considered commercially sensitive in a competitive environment (Llewellyn and others, 2013). However, Monitor are asking trusts to voluntarily share their PLICS data with them for 2013/14, to explore how these might be used in improving tariff price setting.

The significant annual fluctuation in costs reported has resulted in instability of tariff prices year to year, which acts as a barrier to long-term planning, innovative investment and reconfiguration decisions (Monitor, 2012). Furthermore, fluctuations do not reflect how providers believe their costs are actually changing. Providers with the ability to do so are cross-subsidising services. While this enables continuity of service provision, it also enables providers to continue running inefficient services, rather than responding to intended efficiency incentives. This also disadvantages providers who are unable to cross-subsidise in order to maintain loss-making services, for example, those that do not provide the same range of services (Monitor, 2012).

With confidence in Payment by Results falling, providers and commissioners are increasingly negotiating local prices. Some flexibility in the system is desirable, and is permitted under Payment by Results rules if it supports local innovation, service redesign or treatment of an unusual mix of patients. For example, one of the biggest foundation trusts in England has recently announced that they have agreed a contract outside the national tariff for the last two years, under which they have received funding growth that has been invested in out-of-hospital care. The trust claims that not being incentivised to admit patients enables them to invest in other services and work across health sectors

(Calkin, 2013). A potential risk of local pricing arrangement arises if local negotiations are based on poor data or historic prices, and do not reflect current or efficient costs. Where below-tariff prices are negotiated, any decline in quality must be guarded against. Quality monitoring and, increasingly, incentives for quality improvement are being incorporated into secondary care payment systems, with several pay-for-performance schemes introduced in recent years.

Box 3.1: International evidence on diagnosis-related payment

Evidence collected from numerous countries that have adopted diagnosis-related methods of payment suggests intended consequences tend to outweigh unintended consequences. Replacing block budgets with diagnosis-related payments increases productivity and efficiency, at least in the short term, with decreases in length of stay and increases in activity levels seen (Street and others, 2011).

Several Scandinavian countries, which have health financing and delivery arrangements similar to those in England, have adopted diagnostic case-based payment for some or all of their hospitals in the last couple of decades. 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 a year in Sweden as a whole between 1989 and 1994 (OECD, 2003). The 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. It has been estimated that relative efficiency rose by ten per cent over the period from 1989 to 1995 in the counties that adopted diagnosis-related payments (Gerdtham and others, 1999). These relative gains in productivity seem to have derived from modest increases in admissions, combined with sharp falls in length of stay and beds. However, the productivity effects seem to have been temporary, 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 between 1994/95 and 1997/98 in Sweden as a whole. The introduction of activity-based funding had made it more difficult to control total costs and this led some counties to limit activity and in some cases to introduce discounts, or no additional payment, for activity exceeding the ceilings (Kastberg and Siverbo, 2007).

Increased activity is beneficial only as long as additional activity was the intention of the incentive, for example, to reduce waiting times, and one study found increased patient satisfaction associated with reduced waiting times after the introduction of diagnosis-related payment (Hagen and others, 2006). Little evidence exists to show that the quality of care changes significantly following the adoption of diagnosis-related payments, based on studies in a number of countries (Or and Häkkinen, 2011). Effects of introducing these types of payments are, however, context-specific, and will vary depending on what was in place previously, and with changes to concurrent policy and overall funding levels. Table 3.1 summarises the impact of introducing diagnosis-related payment on hospital activity and length of stay. While length of stay has consistently reduced, activity has increased in European countries that have moved from a block budget system, but decreased in the US, where a fee-for-service system preceded diagnosis-related payment.

Table 3.1: Hospital activity and length of stay under diagnosis-related payments

	Country, year of implementation of DRGs	Study	Activity	Average length of stay
US 1980s	US, 1983	US Congress – Office of Technology Assessment, 1985	▼	▼
		Davis and Rhodes, 1988	▼	▼
		Guterman and others, 1988	▼	▼
		Kahn and others, 1990		▼
		Manton and others, 1993	▼	▼
		Muller, 1993	▼	▼
		Rosenberg and Browne, 2001	▼	▼
European countries 1990s/2000s	Sweden, early 1990s	Anell, 2005	▲	▼
		Kastberg and Siverbo, 2007	▲	▼
	Italy, 1995	Louis and others, 1999	▼	▼
		Ettelt and others, 2006	▲	
	Spain, 1996	Ellis and Vidal-Fernández, 2007	▲	
	Norway, 1997	Biøm and others, 2003	▲	
		Kjerstad, 2003	▲	
		Hagen and others, 2006	▲	
		Magnussen and others, 2007	▲	
	Austria, 1997	Theurl and Winner, 2007		▼
	Denmark, 2002	Street and others, 2007	▲	
	Germany, 2003	Böcking and others, 2005	▲	▼
		Schreyögg and others, 2005		▼
		Hensen and others, 2008	▲	▼
	England, 2003/04	Farrar and others, 2007	▲	▼
		Audit Commission, 2008	▲	▼
Farrar and others, 2009		▲	▼	
France, 2004/05	Or, 2009	▲		

Source: Street and others, 2011; quoted in Charlesworth and others, 2012.

The number and extent of policy objectives of DRG payments has varied widely, with Table 3.2, below, highlighting the high ambitions for DRG payments in England compared to other European countries.

Table 3.2: Policy objectives for DRG payments in European countries

	England	Finland	France	Germany	Ireland
Increase efficiency	✓		✓	✓	✓
Expand activity	✓				
Enhance patient choice	✓				
Increase patient satisfaction	✓				
Reduce waiting lists	✓				
Improve quality	✓		✓	✓	
Control costs	✓				
Ensure the fair allocation of resources (or funding) across geographical areas, and across and within health care sectors	✓	✓	✓	✓	
Shift patterns of service provision away from historical patterns	✓				
Encourage the development of new, cost-effective treatment pathways	✓				
Improve transparency of hospital funding, activity and management	✓		✓	✓	✓
Encourage providers to be responsible to patients and purchasers	✓				
Cover costs of production		✓			
Create a level playing field for payments to public and private hospitals			✓		
Improved documentation of internal processes and increased managerial capacity, which would in turn improve efficiency and quality				✓	
Establish link between activity and funding		✓			✓

Source: O'Reilly and others, 2012; quoted in Sangan, 2013.

Pay-for-performance schemes in secondary care

Pay-for-performance schemes in secondary care in England have not translated into incentives for individual staff, only for hospitals as institutions. There has been variation in whether hospitals have transmitted bonuses achieved or penalties incurred to teams or areas of work through their internal budget allocations.

Advancing Quality

Evaluation of this tournament-style scheme found a reduction in mortality for the three incentivised conditions evaluated, estimated to be equivalent to 890 fewer deaths over the 18 months (Sutton and others, 2012). This was only statistically significant for one of the conditions, however, and the differential impact illustrates the importance of selection of performance targets with room for, and feasibility of, improvement, in order to maximise value for money. This improvement in patient outcomes is in contrast to the US scheme on which it was based, with which no improvement was seen. The evaluators suggest this could be due to the fact that Advancing Quality introduced wider quality improvement measures in parallel with the financial incentive, including specialist nurses in hospitals, improved data collection linked to regular feedback about performance to clinical teams, and shared learning meetings between hospitals. The size of bonuses was also greater in the English scheme than in the US scheme, and a greater proportion of hospitals were able to earn bonuses. Preliminary analysis suggests that if only the cost of the bonuses is considered, Advancing Quality was cost-effective. However, other costs incurred in quality improvement measures were not included in this analysis (Sutton and others, 2011).

CQUIN

A national evaluation of the 2010/11 CQUIN schemes (including acute care, ambulance service, community care providers, and mental health and learning disability providers) identified a high level of diversity of schemes, topics and indicators (McDonald and others, 2013). Across 337 schemes, there were 113 distinct goal topics, with a total of 5,001 indicators used (over 3,000 of which were unique). Although the intention was for official national indicators to be used to support local performance goals, more than half (57 per cent) of those used in acute care were locally developed (with just 12 per cent being national, and the rest regional). The schemes were highly complex, with a single scheme in acute care having up to 25 goals and 52 indicators. Although there was relatively high agreement within sectors on local goals, these were measured by many different locally developed indicators. Patient or user satisfaction was the most commonly included goal. Although this was a nationally mandated goal for acute schemes, it also represented over a third of local goals in this sector, and was included in more than three quarters of schemes in other sectors.

It was also intended that CQUIN goals would be both evidence-based and focused on outcomes. Despite this, many of the locally agreed indicators concerned structures and processes, and were based on, at best, weak evidence of effectiveness (McDonald and others, 2013). In addition, baseline data were absent for the majority of indicators, with almost half of indicators in 2010/11 schemes having baseline 'to be confirmed', and a further quarter 'not available'. Setting meaningful indicators requires good evidence of the association between structures, processes and outcomes, and the potential impact of incentivised actions by providers. This not only adds burden at the local level, but has also frequently resulted in unclear or imprecise indicators. This, together with the lack of consistency between schemes, limits the ability to benchmark and compare schemes to evaluate effectiveness of payment approaches. This should be a critical aspect of any scheme introduced, given the scarcity of evidence.

The CQUIN framework was found to have been successful in helping commissioners and providers to jointly identify and prioritise local needs for quality improvement. There are reports that relationships between commissioners and providers have been strengthened through the process of agreeing CQUINs, although the involvement of front-line

clinicians was often found to be lacking, with managers participating in negotiations (McDonald and others, 2013).

The evaluation found that while the local, 'bottom-up' approach worked well in identifying relevant goals for local pay-for-performance schemes, inclusion of particular goals did not lead to statistically significant performance improvements in related outcome indicators. It is suggested that the technical design of the schemes – which, including local proliferation of indicators, was not consistent with the original intentions of the Department of Health – could explain this lack of effect. The evaluators conclude that while there is an important case for local strategic and clinical input, this should be separated from the complex technical processes of defining indicators, agreeing thresholds and setting prices. The necessary expertise is unlikely to be available locally in all areas, and even if it was, repetition of technical design would not be efficient. They suggest a firmer national framework would be beneficial, perhaps with a 'menu' of national indicators that commissioners and providers could select from once they had identified priorities (McDonald and others, 2013). Standardisation of indicators would also benefit evaluation.

Best practice tariffs

Widespread support was found for BPTs among clinicians, who tended to prefer BPTs to CQUINs. They viewed BPTs as being more evidence-based and having fairer payment structures, and welcomed the clarity provided by their top-down design. Commissioners, in contrast, were more familiar with CQUIN (McDonald and others, 2012). Around half of acute trusts responding to an evaluation survey had taken up the BPT for stroke and hip fracture, but only around a quarter for cataracts. Barriers to uptake included the burden of changing data collection, the fact that the tariff did not reflect the reality of multiple conditions, and that it was not seen as incentivising major changes in practice (McDonald and others, 2012). This illustrates well the importance of involving clinicians in the design of schemes, and of understanding current practice and other parallel initiatives.

For the BPT incentivising day-case treatment for cholecystectomy, which applied to all trusts, the price offered for cases planned and treated as day cases was increased by 24 per cent, while the rate for inpatient treatment remained unchanged. For the first year of the BPT, the day-case rate increased by seven per cent, with no evidence of providers selecting patients more amenable to day-case treatment, reduced quality, or increased total volume of treatment. However, this change was accompanied by an increased waiting time for treatment, of 14 days, illustrating the need to ensure resources match incentivised changes (McDonald and others, 2012). Of hospitals receiving the BPT for stroke, this accounted for just over half of episodes. There was no evidence of an impact on quality or outcomes over and above improvements achieved nationally through additional activities to improve quality of stroke care (McDonald and others, 2012). This again emphasises that pay-for-performance scheme priorities should be selected to target areas where there is evidence to suggest room for quality improvement, as schemes are unlikely to be cost-effective in areas where other types of quality improvement initiatives have already been implemented successfully. This also highlights the need to compare both the costs and benefits, and the level of synergism, of payment and non-payment quality improvement interventions.

In hospitals receiving the BPT for hip fracture, two fifths of episodes received the BPT. In this case, improvements were found in both process quality and outcomes. Although the proportion of patients receiving surgery within 48 hours of admission was rising nationally in 2010/11, the increase was greater in hospitals receiving the BPT (by four

per cent). There was also a greater decrease in the mortality rate (by 0.7 per cent in BPT receivers versus non-receivers) and a greater increase in the proportion of patients discharged to their usual place of residence within 56 days (2.1 per cent greater). The evaluators suggest that the difference in impact between BPTs could be due to the structure of the tariffs (the hip fracture BPT is only paid if all criteria are met), differences in underlying quality trends, and parallel quality improvement initiatives (McDonald and others, 2012).

Block contracts in secondary care

While a large part of acute and emergency care is paid for through activity-based methods, mental health and community health services are primarily reimbursed through block contracts. Without a link to quality and activity, this will not deliver incentives to improve quality or efficiency. Indeed, research suggests mental health services have significant scope to improve their efficiency, with very large variations in activity levels (Naylor and Bell, 2010). There is little data on community care, but what exist suggest there is also huge potential for productivity improvements in the sector, with observational research suggesting patient-facing time could be increased by 25 to 30 per cent (NHS Institute for Innovation and Improvement, 2009).

Block contracts can be managed in such a way as to incentivise quality and efficiency. However, the development of incentives and effective pricing of services require robust, accurate and up-to-date information about costs of treating individuals, in order to create effective and efficient incentives. The quality of cost information for mental health and community services is poor, and reported costs vary widely between providers (Monitor, 2012). This will need to be urgently addressed to support attempts to expand activity-based payment in these sectors. In recognition of this, mandatory submission of a minimum dataset for mental health services has been introduced, and will be brought in for community health services from 2014/15. The introduction of activity-based payment in the acute sector vastly improved the quality of data collection, coding and reporting. Consideration will need to be given to how to achieve these improvements in the mental health and community sectors in the absence of activity-based payment.

Summary of evidence

The evidence from the evaluations of payment approaches supports a role for these in improving quality and productivity. There is a lack of evidence to date to suggest they can confidently be used in incentivising better patient outcomes.

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 simply what is measurable. This will foster confidence in the payment system, which is currently lacking, resulting in reduced adherence and blunting of the intended incentives.

There are roles both for central and local bodies. While local identification of goals, which should involve clinicians, is beneficial in many ways, technical design is better performed centrally, due to the expertise needed and the burden it adds to local areas. Consistency in design and indicators provided by central determination will also enable benchmarking and comparisons, and hence build the limited evidence base for the role and design of payment systems.

4. Summary

The payment system is a lever to support the delivery of high-quality, cost-effective care, but it is only one lever among multiple available to commissioners. It is imperative that the limitations of a payment approach are recognised in the objectives set for it, which should be based on what good evidence tells us payment approaches may be able to achieve. The objectives that have been set for the redesigned payment system for NHS care are highly ambitious. There are intentions that it will reimburse outcomes, support changing patterns of care and better links between health and wider sectors, as well as enable the efficient allocation of resources. These are perhaps overly ambitious in light of what evidence tells us the payment approaches can achieve.

The evidence presented in this report supports a role for financial incentives in improving the quality and productivity of processes of care. However, there is a lack of evidence for an impact on patient outcomes. It is important therefore that a distinction is made between primary objectives for the payment system, based on robust evidence for an effect, and those factors that the payment system should enable and support. Based on the research, primary objectives supported by evidence are as follows:

- to incentivise improvements in quality
- to incentivise improvements in efficiency and productivity
- to ensure resources are allocated both appropriately and efficiently, following the patient and matching need rather than demand
- to ensure transparency and accountability for the use of public resources.

In addition to these primary objectives, the payment system must enable and support other aims of the health service. There is a lack of evidence for a primary role of payment in directly incentivising these, and non-payment approaches may be more effective in achieving changes in these other areas. At the very least, it must be ensured that the payment system is not acting as a barrier to achieving change in these areas, through unintended consequences of approaches taken. These wider ambitions of the health service that the payment system must support include:

- achievement of outcomes
- better integration and coordination of services both within and between sectors
- patient choice
- innovation, both in health care and health care payment systems.

What is clear from the evidence is that better data, better use of data, and better evaluation of payment systems must be priorities. Until this can be made a reality, there is a case for caution regarding what payment reform can achieve. The most important opportunities in the short to medium term are around ensuring that the payment system allocates resources efficiently and does not create perverse incentives or undermine the other policy levers that can influence outcomes for patients. This urgently requires better

activity and cost data across sectors, particularly in community, mental health and primary care services. Without this, cross-system goals will not be possible.

The roles of local and central bodies also must be defined. While local identification of goals is beneficial in many ways, there is a key role for central bodies in providing resources for coding, pricing and indicator setting where replication of these tasks at local level would consume too many resources or create confusing variation.

As Monitor and NHS England develop the NHS payment system, it is important that any immediate changes made to the payment system are in line with the direction of longer-term changes. Stability is important to both providers and commissioners, and for compliance with the payment system. A long-term direction should be clearly signalled through the payment system, however, this approach must also recognise that unforeseen changes in best practice, technologies or priorities will require some flexibility.

Our accompanying policy response, *NHS payment reform: lessons from the past and directions for the future*, makes policy recommendations to Monitor and NHS England for the reform of the payment system, taking into account the evidence reviewed here.

References

- Calkin S (2013) '£560m foundation trust dumps payment by results', *Health Service Journal*. www.hsj.co.uk/news/acute-care/560m-foundation-trust-dumps-payment-by-results/5059535.article
- Charlesworth A, Davies A and Dixon J (2012) *Reforming Payment for Health Care in Europe to Achieve Better Value*. Nuffield Trust.
- Department of Health (2005) *Health Reform in England: Update and next steps*.
- Department of Health (2008) *Transforming Community Services: Currency and pricing options for community services*.
- Department of Health (2009) *The Operating Framework for the NHS in England 2010/11*.
- Department of Health (2010) *The Operating Framework for the NHS in England, 2011/12*.
- Department of Health (2011a) *A Simple Guide to Payment by Results*.
- Department of Health (2011b) 'More choice for NHS patients across the country', Press release, 19 July, http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/MediaCentre/Pressreleases/DH_128538
- Department of Health (2011c) *The Operating Framework for the NHS in England, 2012/13*.
- Department of Health (2012a) *A Simple Guide to Payment by Results*.
- Department of Health (2012b) *Code of Conduct for Payment by Results in 2012-13*.
- Department of Health (2012c) *Mental Health PbR Road Test Package for 2013-14 Guidance*.
- Department of Health (2012d) *The Mandate: A mandate from the government to the NHS Commissioning Board*.
- Department of Health (2012e) *QIPP Long Term Conditions. Supporting the local implementation of the year of care funding model for people with long term conditions*.
- Department of Health (2013a) *Annual Report and Accounts 2012/13*.
- Department of Health (2013b) *Payment by Results Guidance for 2013/14*.
- Dixon A, Khachatryan A, Wallace A, Peckham S, Boyce T and Gillam S (2010) *The Quality and Outcomes Framework: Does it reduce health inequalities? Final report*. NIHR Service Delivery and Organisation programme. www.netscc.ac.uk/hsdr/files/project/SDO_ES_08-1716-207_V01.pdf
- Doran T and Roland M (2010) 'Lessons from major initiatives to improve primary care in the United Kingdom', *Health Affairs* 29(5), 1023–1029.
- Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C and Reeves D (2011) 'Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework', *BMJ* 342, d3590.
- Dredge R (2004) *Hospital Global Budgeting*. Washington: International Bank for Reconstruction and Development/World Bank.
- Ellis RP and McGuire TG (1990) 'Optimal Payment Systems for Health Services', *Journal of Health Economics* 9(4), 375–396.
- Ellis RP and McGuire TG (1993) 'Supply-side and Demand-side Cost Sharing in Health Care', *Journal of Economic Perspectives* 7, 135–151.
- 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*. Report to the Department of Health.
- Farrar S, Sussex J, Yi D, Sutton M, Chalkley M, Scott T and Ma A (2007) *National Evaluation of Payment by Results*, Report to the Department of Health.
- Farrar S, Yi D, Sutton M, Chalkley M, Sussex J and Scott A (2009) 'Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis', *British Medical Journal* 339, b3047.
- Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E and Beyer FR (2011) *An Overview of Reviews Evaluating the Effectiveness of Financial Incentives in Changing Healthcare Professional Behaviours and Patient Outcomes*. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009255/abstract>

- Geissler A, Quentin W, Scheller-Kreinsen D and Busse R 'Introduction to DRGs in Europe: common objectives across different hospital systems', Chapter 2 in Busse R, Geissler A, Quentin W and Wiley M (2011) *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. European Observatory on Health Systems and Policies Series.
- Gerdtham U, Löthgren M, Tambour M and Rehnberg C (1999) 'Internal markets and efficiency in health care: a multiple output stochastic frontier analysis', *Health Economics* 8, 151–164.
- Gillam S (2013) 'The Quality and Outcomes Framework: where next?', *British Medical Journal* 346, f659.
- Glasziou PP, Buchan H, Del Mar C, Doust J, Harris M, Knight R, Scott A, Scott IA and Stockwell A (2012) 'When financial incentives do more good than harm: a checklist', *British Medical Journal* 345, e5047.
- Hagen TP, Veenstra M, Stavem K (2006) *Efficiency and Patient Satisfaction in Norwegian Hospitals*. Health Economics Research Programme at the University of Oslo, Working Paper 2006, 2.
- Health & Social Care Information Centre (2013) *Do QOF Achievement Scores Shown for PMS Practices Incorporate a PMS Deduction?*
- Health Foundation (2011) *Do Quality Improvements in Primary Care Reduce Secondary Care Costs?*
- Health Select Committee (2013) *Public Expenditure on Health and Care Services: Eleventh report of session 2012–13* (HC 651).
- HM Treasury (2013) *Budget 2013* (HC1033). The Stationery Office.
- Holmstrom B and Milgrom P (1991) 'Multitask principal-agent analyses: incentive contracts, asset ownership and job design', *Journal of Law, Economics and Organization* 7 (special issue), 24–52.
- Kastberg G and Siverbo S (2007) 'Activity-based financing of health care – experiences from Sweden', *Int J Health Plann Mgmt* 22, 25–44.
- Llewellyn S, Chambers N, Ellwood S, Whitfield T, Yarwood C, Adil M, Begkos C and Wood C (2013) *Patient level costing and information systems (PLICS): Does PLICS enable cost savings, service integration and better understanding of cost drivers in healthcare systems*. Abstract presented at EHMA Annual Conference 2013. www.ehma.org/files/AB%20Website.pdf
- Maisey S, Steel N, Marsh R, Gillam S, Fleetcroft R and Howe A (2008) 'Effects of payment for performance in primary care: qualitative interview study', *J Health Serv Res Policy* 13, 133–139.
- Mannion R and Davies HTO (2008) 'Payment for performance in health care', *British Medical Journal* 336, 306–308.
- Marini G and Street A (2007) 'A transaction costs analysis of changing contractual relations in the English NHS', *Health Policy* 83, 17–26.
- McDonald R, Zaidi S, Todd S, Konteh F, Hussain K, Roe J, Allen T, Fichera E and Sutton M (2012) *A Qualitative and Quantitative Evaluation of the Introduction of Best Practice Tariffs. An evaluation report commissioned by the Department of Health*. University of Nottingham.
- McDonald R, Zaidi S, Todd S, Konteh F, Hussain K, Brown S, Kristensen SR and Sutton M (2013) *Evaluation of the Commissioning for Quality and Innovation Framework Final Report*. University of Nottingham.
- Mikkola H, Keskimäki I and Häkkinen U (2001) 'DRG-related prices applied in a public health care system – can Finland learn from Norway and Sweden?', *Health Policy* 59, 37–51.
- Monitor (2012) *An Evaluation of the Reimbursement System for NHS-funded Care*.
- Monitor and NHS England (2013a) *How Can the NHS Payment System Do More for Patients?*
- Monitor and NHS England (2013b) *The National Tariff 2014/15*.
- National Audit Office (2008) *NHS Pay Modernisation: New contracts for general practice services in England*. Report by the Comptroller and Auditor General, HC 307, Session 2007–2008.

- Naylor C and Bell A (2010) *Mental Health and the Productivity Challenge: Improving quality and value for money*. The King's Fund.
- Naylor C, Imison C, Addicott R, Buck D, Goodwin N, Harrison T, Ross S, Sonola L, Tian Y and Curry N (2013) *Transforming our health care system: Ten priorities for commissioners*. The King's Fund.
- NHS Confederation 'How sector regulation will work'. www.nhsconfed.org/priorities/NHS-reforms/Regulation/Pages/sector-regulation.aspx
- NHS England (2012) *Quality Premium: 2013/14 guidance for CCGs. Draft*.
- NHS England (2013a) *NHS Authorities and Trusts*.
- NHS England (2013b) *Putting Patients First: The NHS England business plan for 2013/14 – 2015/16*.
- NHS England (2013c) *Commissioning for Quality and Innovation (CQUIN): 2014/15 guidance*.
- NHS Institute for Innovation and Improvement (2009) *Productive Community Services: Releasing time to care*.
- NHS Supply2Health *AQP Pricing Principles*.
- OECD (2003) *Health at a Glance: OECD Indicators 2003*.
- Or Z and Häkkinen U 'DRGs and quality: for better or worse?' Chapter 8 in Busse R, Geissler A, Quentin W and Wiley M (2011) *Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. European Observatory on Health Systems and Policies Series.
- Pike C (2010) *An Empirical Analysis of the Effects of GP Competition*. Co-operation and Competition Panel Working Paper Series, Volume 1, Number 2.
- Quentin W, Scheller-Kreinsen D and Busse R 'Technological innovation in DRG-based hospital payment across Europe', Chapter 9 in Busse R, Geissler A, Quentin W and Wiley M (2011) *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. European Observatory on Health Systems and Policies Series.
- Roberts A, Marshall L and Charlesworth A (2012) *A Decade of Austerity? The funding pressures facing the NHS from 2010/11 to 2021/22*. Nuffield Trust.
- Sangan (2013) *Lessons from Europe: The value of tariff redesign*, Health Services Research Network Briefing 16; original source O'Reilly and others (2012) 'Paying for hospital care: the experience with implementing activity-based funding in five European countries', *Health Economics, Policy and Law*.
- Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J and Young D (2011) *The Effect of Financial Incentives on the Quality of Health Care Provided by Primary Care Physicians*. Cochrane Database Systematic Review, 7(9), CD008451.
- Shaw S, Rosen R and Rumbold B (2011) *What is Integrated Care?* Nuffield Trust.
- Street A, O'Reilly J, Ward P and Mason A 'DRG-based hospital payment and efficiency: theory, evidence and challenges', Chapter 7 in Busse R, Geissler A, Quentin W and Wiley M (2011) *Diagnosis-related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. European Observatory on Health Systems and Policies Series.
- Sutton M, Nikolova S and McDonald R (2011) 'Impact of a pay-for-quality programme in the North West of England on patient mortality risk', paper presented to the Health Economists' Study Group, York.
- Sutton M, Nikolova S, Boaden R, Lester H, McDonald R and Roland M (2012) 'Reduced mortality with hospital pay-for-performance in England', *New England Journal of Medicine* 367, 1821–1828.
- Tambour M and Rehnberg C (1997) *Internal Markets and Performance in Swedish Health Care*, Stockholm School of Economics, The Economic Research Institute, Working Paper No.161.

About the authors

Louise Marshall

Louise Marshall is a Specialty Registrar in Public Health, who previously held a trainee position at the Nuffield Trust. Prior to working at 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.

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.

Jeremy Hurst

Jeremy Hurst is an independent health economist. He has been a senior economic adviser at the Department of Health, and was head of the former Health Policy Unit at the Organisation for Economic Co-operation and Development (OECD). While he was at the OECD, he contributed to reviews of the Finnish, Korean and Turkish health systems, and to a number of reports and working papers on topics such as performance measures, waiting times, health system reforms, human resources for health care and prevention.

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

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

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

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

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

59 New Cavendish Street
London W1G 7LP
Telephone: 020 7631 8450
Facsimile: 020 7631 8451
Email: info@nuffieldtrust.org.uk

 www.nuffieldtrust.org.uk

Published by the Nuffield Trust.
© Nuffield Trust 2014. Not to be reproduced
without permission.

ISBN: 978-1-905030-76-7