Paying for improvements in quality: recent experience in the NHS in England

RACHEL MEACOCK^{1,*}, SØREN RUD KRISTENSEN¹, MATT SUTTON¹

¹ University of Manchester, UK

Abstract: There is a long-term international trend towards linking payments more closely to providers' performance. The US and England have been at the forefront of the design and introduction of such pay-for-performance (P4P) schemes. England's experience is, however, likely to have greater salience for the Nordic countries' health care systems due to the publicly funded finance structure. We review the development of five of England's major schemes and summarise the available evidence on their impacts. These schemes are: the *Quality and Outcomes Framework* (QOF); *Advancing Quality*; the *Commissioning for Quality and Innovation* (CQUIN) framework; *Best Practice Tariffs*; and the newest 'non-payment' policies. Much of the evidence is limited by the non-experimental way in which the schemes have been introduced, with limited data available prior to the introduction of the schemes and no experimentally unexposed providers to serve as controls. Nonetheless, the existing evidence suggests that P4P can result in modest short-term improvements in the incentivised aspects of performance. There is little evidence of effort diversion, yet some to suggest positive spillovers of these schemes onto non-incentivised aspects of performance. While there is some evidence of gaming and inequitable consequences, these do not appear to be widespread. The gains that can accrue across large patient populations as a result of relatively small financial incentives mean that P4P schemes can be cost-effective. P4P programmes are likely to be most effective when introduced as a supporting part to a wider quality improvement initiative, and when results are published to encourage a reputational as well as a financial incentive for improvement. Though the accumulation of evidence to support P4P has not been systematic or especially robust, it remains a popular policy tool with decision-makers in England, with its reach set to increase further in the future. [291 words]

JEL codes: I1, I18

Key words: pay-for-performance, quality, financial incentives

1 Introduction

Pay-for-performance (P4P) schemes link financial payments made by purchasers explicitly to the quality of care delivered by health care providers. Quality is generally assessed using measures of the clinical processes judged to represent best treatment practices. It is theorised that improvements on these quality metrics will in turn lead to better patient outcomes. Such schemes have been introduced to overcome problems with the existing fee-for-service (FFS) financing models, which incentivised providers to

* **Correspondence to:** Rachel Meacock, Manchester Centre for Health Economics, Institute of Population Health, The University of Manchester, 4.312 Jean McFarlane Building, Oxford Road, Manchester M13 9PL, UK. Phone: + 44 1612757685, E-mail: rachel.meacock@manchester.ac.uk

maximise the quantity of services provided whilst minimising costs (Greene and Nash, 2009). Reducing variation in clinical practice has also been given as justification for their use (Maynard, 2012).

The past decade has seen a growing adoption of P4P programmes in both primary and secondary care, particularly in the United States and England. So far, the Nordic experience with P4P is limited, but there is a considerable interest among policy makers in the provision of incentives for high quality care. Small-scale initiatives for linking reimbursement to quality indicators are currently being considered and tested, but large national programmes have yet to be implemented.

Sweden implemented a national P4P scheme in 2008, incentivising performance on the national waiting time targets at the county council level. Swedish county councils are also implementing P4P schemes locally, mainly aimed at improving quality in primary care, and to a lesser extent in secondary care (Anell et al., 2012).

In Denmark, following a debate on the appropriateness of incentives embedded in the activity based reimbursement scheme for hospitals, the Danish Government formed a “committee for better incentives in the health care system” in 2012. One year later, a report was published considering how financial incentives for higher quality care could potentially be provided (Ministeriet for Sundhed og Forebyggelse, 2013). The report was cautious about a national P4P scheme because of perceived barriers in the current reimbursement model, and due to fears of possible unintended effects resulting from focusing rewards upon specific dimensions of quality. It did, however, encourage regional experiments with P4P, which are now being developed and piloted in three of the five regions.

The Norwegian Ministry of Health and Care Services began work on how to improve quality and patient safety in the Norwegian health care system in 2011, and as part of this work published a report on international experiences with P4P (Helsedirektoratet, 2013a). From 1st January 2014, Norway will launch a three-year P4P initiative distributing 500m NOK to the 4 Health Regions responsible for hospital care, on the basis of the regions’ performance on indicators of process, outcome and patient-experienced quality (Helsedirektoratet, 2013b).

We are not aware of any P4P schemes that have been introduced or are planned in Finland. As P4P initiatives are gradually emerging in most Nordic countries, England possesses a decade of experience in designing and implementing such schemes. Previous reviews have summarised the published international evidence on the effects of P4P programmes (Eijkenaar et al., 2013). We propose that England’s experiences offer significant learning potential for the Nordic countries, at a time before large national schemes have been rolled out. With its Beveridge-style health care system, the lessons that can be drawn from the English experience are likely to be more relevant to the Nordic countries than those of countries such as the US, which run under very different institutional arrangements.

In this paper we aim to summarise the development of P4P schemes in England, with particular reference to the five largest programmes introduced to date. We first describe the main features of each policy and the time-line of their introduction, before presenting the available evidence on their consequences. Some general lessons are then drawn from England’s experience, and the future of P4P discussed.

2 Pay-for-performance in the NHS in England

P4P schemes can be classified in terms of their design features, and many different approaches have been used in the programmes introduced to date. Some key design aspects include:

- the level of provider to be targeted (individual health care professionals, clinical teams or whole hospitals)
- whether participation by providers is mandatory or voluntary
- the patient groups to be covered
- the aspects of quality to be incentivised (data collection, process, outcome or patient experience measures)
- whether the payments represent bonuses or penalties to providers
- the structure of the payment schedules (linear payment schedules (with or without minimum or maximum thresholds) or target payments)
- whether rewards are given on absolute, relative, or improvements in, performance
- the size of the bonuses available
- the source of funding from the payer's perspective (additional money vs. reallocation of current budgets)
- how performance is monitored (self-reporting vs. independent data collection)
- the frequency of assessment
- the supporting levers that accompany the scheme (public reporting, feedback and/or shared learning)

A variety of P4P schemes have been introduced by the NHS in England, an overview of which are summarised in Table 1. The key features of these programmes are described in turn.

2.1 The Quality and Outcomes Framework

The Quality and Outcomes Framework (QOF), introduced for general practices (primary care) in 2004, was the first major P4P scheme to be implemented (Gravelle et al., 2010; Roland, 2004). It was introduced in all four countries of the UK and, while voluntary in principle, had close to universal participation. Each practice was rewarded for its performance on 146 quality indicators covering four domains: clinical management of 10 chronic conditions, practice organisation, patient experience, and additional services offered. A maximum of 1050 'points' could be scored by a practice across these indicators, with the average practice, containing four general practitioners (GPs), able to gain roughly £130,000 in 2005/6 if they achieved the maximum points level. Rewards were based on absolute performance, and structured using a linear payment schedule with a minimum threshold of 25% below which no payments were made, and maximum thresholds between 50% and 90% above which no additional payments were made. Approximately £1billion was paid out to practices in 2005/6, financed using additional funding from the expanding NHS budget. GPs retained this money as personal income, which was paid out annually based on self-reported performance data. These performance data are published on the publicly accessible Health and Social Care Information Centre and NHS Choices websites, and are subject to further scrutiny at local level by purchasers and other providers.

Changes to the QOF are negotiated annually by the doctors' trade union (the British Medical Association) and the Department of Health, based on technical advice provided by the National Institute for Health and Care Excellence (NICE). There has been substantial revision of the scheme during the ten years of its existence. These include the

Table 1: An overview of pay-for-performance policies introduced in England

	QOF	AQ	CQUIN	BPT	NPP
Year of introduction	2004	2008	2009	2010	2009
Targeted providers	Primary care practices	Hospital clinical teams	Hospitals	Hospitals	Hospitals
Participation	Voluntary	Voluntary	Mandatory	Mixed	Mandatory
Incentive	Bonuses	Bonuses	Penalties	Mixed	Penalties
Form of reward	Income	Budget	Revenue	Revenue	Revenue
Payment schedule structure	Linear	Tournament	Target payment	Per-patient	Per-patient
Rewards based on	Absolute performance	Relative performance	Absolute performance	Absolute performance	Absolute performance
Size of incentive	25% of total income	2 - 4% of tariff	0.5 - 2.5% of contract	5 - 43% of tariff	100% of tariff
Monitoring	Self-reported	Self-reported	Mixed	Admin data	Admin data
Frequency of payment	Annual	Annual	Annual	Continuous	Continuous
Supporting levers	Public reporting	Public reporting Provider engagement Shared learning	None	None	None

removal and addition of indicators, revisions of thresholds, re-allocations of points across indicators, and changes in the amounts paid per point and the formula used to adjust the amount each practice is paid per point according to its level of disease prevalence. In the latest round of negotiations, the number of indicators has been substantially reduced and the released funding will be paid according to a capitation formula rather than by performance.

2.2 Advancing Quality

Advancing Quality (AQ) was the first hospital-based P4P scheme in the UK, introduced in just one of the 10 English health regions. Modelled on the US Premier Hospital Quality Incentive Demonstration (HQID), AQ covered all 24 NHS hospitals in the North West region of England (population 6.8 million = 13% of the English population). Performance on 28 quality indicators was targeted, covering patients admitted in an emergency for three health conditions (acute myocardial infarction (AMI), heart failure, pneumonia) and for two types of planned surgery (coronary artery bypass grafting (CABG) and hip and knee replacements). The scheme ran as a stand-alone initiative from the 1st October 2008 to 31st March 2010, after which it was absorbed into the national Commissioning for Quality and Innovation (CQUIN) programme (see section 2.3).

For the first 12 months the scheme ran as a pure tournament, with the top quartile of performers on each condition (on an aggregated measure across all indicators) receiving a bonus payment equal to 4% of the national tariff revenue received for that activity, and

those in the second quartile a 2% payment. For the remaining 6 months of the initiative, the bonus system changed so that it was possible to earn rewards based on three criteria. An attainment bonus was awarded to a hospital if its achievement during these 6 months exceeded the median score across all hospitals in the first year. If this attainment standard was reached a hospital also became eligible for two further payments. The improvement bonus was awarded if a hospital's improvement in achievement compared to the first 12 months scored in the top quartile of improvements, and the achievement bonus was given to hospitals scoring in the top two quartiles of absolute achievement during this 6 month period. A total of £4.8m in incentive payments were paid out during the 18 months of the policy, and awarded to the clinical teams that won them to reinvest into their services. AQ was financed by a reallocation of the North West commissioning budget, so no additional funds were assigned to the policy. Although participation was voluntary, the initiative achieved universal uptake by providers in the region. In addition to the financial incentives, providers participated in shared learning activities as part of the AQ initiative. These included meetings where the top performers would share their tips for success, creating a collaborative rather than competitive environment.

2.3 Commissioning for Quality and Innovation

The CQUIN framework was introduced across England in April 2009 (Department of Health, 2008a, 2008b). The scheme covers providers of acute, ambulance, community, mental health and learning disability services, making a proportion of their total annual income conditional upon the achievement of specific goals for quality and innovation. The CQUIN framework can therefore be seen as a penalties rather than a bonuses regime, with expected funds being withheld if agreed performance levels are not met.

The indicators included in CQUIN are agreed locally between commissioners and providers, which distinguishes it from previous P4P initiatives. Policy makers believed that this local aspect would generate more enthusiasm from providers, ensure an appropriate level of 'stretch' given providers' initial performance and therefore expose them to less financial risk, and safeguard against gaming and an unintended distortion of resources away from areas not targeted by the scheme (Department of Health, 2008c).

As part of the national policy framework, guidelines (Department of Health, 2008a) were published outlining the dimensions of care that locally designed CQUIN schemes should cover, the size of the financial incentives to be linked to performance, and guidance on the choice of indicators to be selected. These guidelines specify that all local CQUIN schemes must incentivize improvement in each of the four quality dimensions: safety, effectiveness, patient experience, and innovation. Local scheme designers are encouraged to select outcome measures over process and structure indicators, and to use existing national indicators where available.

The size of the incentive linked to CQUIN targets increased from 0.5% of a provider's annual contract value in 2009/2010 to 1.5% in 2010/2011 and 2.5% from April 2012, making the CQUIN scheme comparable to the QOF in terms of the overall financial value of potential bonuses. With a fixed value of the programme and no limits on the number of indicators to be incentivised, local scheme designers have the option of negotiating a few high powered goals or many lower powered ones. The relative importance of each indicator is reflected by an assigned weight, which represents the proportion of the total CQUIN payment attached to that measure. While CQUIN rests on an emphasis on local negotiations and ownership, two national goals were introduced for acute providers in 2010/11, and other national goals have since followed.

2.4 Best Practice Tariffs

In April 2010, the Department of Health in England introduced Best Practice Tariffs (BPTs) to the activity-based financing system, with the aim of reducing unexplained variation in care and universalising best practice (Department of Health, 2008b). The scheme initially linked financial rewards to service delivery across four high-volume clinical areas: cataract surgery, gall bladder removal, stroke, and fragility hip fracture. A separate model of best practice was developed for each of the clinical areas, using evidence on the most clinically- and cost-effective way to deliver care for that condition. The scheme for gall bladder removal was applied automatically in the national payment system and therefore applied to all hospitals providing NHS care in England. Participation in the three remaining BPT schemes on the purchaser and provider side was left for local negotiation.

Participating hospitals were rewarded financially for delivering care according to the best practice models. For example, the gall bladder removal BPT encouraged the surgery to be performed as a day case procedure where appropriate. This was incentivised by increasing the price offered to providers for gall bladder removal both planned and performed as a day case by 24%, whilst leaving the amount reimbursed for inpatient treatment unchanged. Payments were made on a per-case basis and on absolute performance on the specified quality measures, and therefore structured as a linear payment schedule with no minimum or maximum thresholds.

The initiative has been expanded each year since its introduction, with BPTs designed to meet one of three aims. Many are focused on changing the setting of care, often from inpatient to day case, as was the case for gall bladder removal. Others aim to streamline care pathways, for example by reducing the amount of outpatient appointments for each patient for which hospitals are reimbursed. Finally, BPTs aim to incentivise the provision of high quality care based upon the best available evidence, as with the provision of designated stroke units.

A further five clinical areas were added to the policy in April 2011, as well as the addition of an extra 12 daycase procedure models. The price differentials were also increased for stroke and hip fragility fracture, providing a larger financial incentive to conform to the best practice models in these areas. In 2012, the policy was expanded further to cover outpatient procedures, same day emergency care, and major trauma. Price differentials were again increased for stroke and hip fragility fracture, and additions made to many of the other existing BPT models. A further six clinical areas were added to the scheme in April 2013 (Department of Health, 2013).

Some of the BPTs were introduced in areas where there was already an on going national clinical audit (e.g. hip fracture) as a supporting lever. The initial intention was to set the price levels for achieving, and failing to achieve, best practice so that providers could only earn additional income compared to the previous year if they reached a pre-specified performance level (e.g. a 60% day case rate). Performance above this level would earn a bonus, and below would represent a penalty compared to the previous year's payment schedule.

2.5 Non-payment policies

The most recent trend in financial incentives, termed non-payment for performance (Rosenthal, 2007) focuses instead on behaviour that payers wish to disincentivise. These policies involve withholding payments when such behaviours occur, rather than paying out bonuses.

2.5.1 Non-payment for Never Events

England first introduced a non-payment policy for so-called ‘Never Events’ in 2009. A Never Event has been defined by the National Patient Safety Agency (NPSA) as “[a] serious, largely preventable patient safety incident that should not occur if the available preventative measures have been implemented by healthcare providers.” (NPSA, 2009). The list of Never Events is updated regularly, and for 2012/13 contained 25 events including wrong site surgery, severe scalding of patients, and unintended retention of a foreign object in a patient after surgical intervention (Department of Health, 2012a).

If a Never Event occurs, providers must initiate an investigation into the causes of the event. In addition, the provider is not reimbursed for both the episode of care that involved the event, and for the costs of consequential treatment (Department of Health, 2012b).

2.5.2 Non-payment for readmissions

In April 2011, England introduced a policy (Department of Health, 2011) according to which hospitals would no longer be reimbursed for emergency readmissions occurring within 30 days of discharge from an elective admission. Around 40% of all readmissions, including those for children under four years of age, maternity, childbirth and cancer patients, and those who self-discharge against clinical advice, were however excluded from these non-payment rules.

The policy was expanded after its first year of operation, and now applies to both emergency and elective first admissions. However, instead of a national rule of non-payment for readmissions, providers and purchasers must now jointly undertake a detailed local clinical review of all readmissions occurring within a fixed review period of between 1 week and 3 months (set by each hospital and commissioner), to determine whether each was avoidable or not. The rate of readmissions deemed avoidable in the review period is then applied to each hospital’s total rate of readmissions for the full financial year, and hospitals will not be reimbursed for readmissions above this locally set benchmark (Department of Health, 2012c). Any savings made by commissioners due to non-payment for readmissions must be reinvested in post-discharge reablement services which support rehabilitation, reablement, and the prevention of future readmissions.

3 Evidence

The widespread adoption of P4P in England has generated much interest in the evaluation of such schemes. Evaluations have focused on the various potential impacts of P4P programmes, both intended and unintended. Studies have examined the impact of policies on the incentivised quality metrics, outcomes for patients treated with conditions covered by the schemes, patient selection, gaming of the system, effort diversion away from non-incentivised treatment areas, spillovers into non-incentivised areas of care, the financial impact of P4P programmes, and their cost-effectiveness.

The amount and quality of evidence available varies by scheme, with fewer results available for those introduced most recently. Below we present the existing research on the consequences of the P4P schemes previously outlined, highlighting any gaps in the current evidence base.

3.1 The Quality and Outcomes Framework

The effects of the QOF have been the subject of much debate. It is difficult, however, to draw definitive conclusions regarding the impact of the policy due to its universal introduction and the coincident introduction of a new data collection system for

monitoring performance. These features result in an absence of control practices not exposed to the intervention, and limited comparable data on the trends in quality prior to the scheme. Despite these barriers to evaluation, a recent systematic review identified 94 articles examining the impact of the initiative (Gillam et al., 2012). These were classified as focusing on five distinct areas: effectiveness, efficiency, equity, patient experience, and professionals' experience and team working.

Studies examining the effectiveness of the scheme suggest that the introduction of the QOF initially lead to improved performance on the incentivised measures, but that many of these indicators may have returned to their long-term trends within three years. No negative short-term effect on non-incentivised indicators was detected, although there is some evidence that these indicators were below their long-term trends after three years. In general, the existing papers that have looked for evidence of spillovers from the incentivised to non-incentivised activities have not considered different forms of spillover explicitly. Sutton et al. (2010) considered two forms: horizontal spillovers, the effect on non-incentivised activities for the patients targeted by the scheme; and vertical spillovers, the effect for untargeted patients on the activities incentivised for the targeted patients. These effects are likely to differ because the mechanisms through which they might arise are different. Horizontal spillovers can arise through more comprehensive care in single consultations (e.g. interventions that address multiple lifestyles) and vertical spillovers can arise if practices introduce new systems for particular activities (e.g. recall systems for blood pressure checks) that are applied across the practice population. Sutton and colleagues found evidence of substantial positive horizontal spillovers, suggesting that the targeting of particular patient groups was the key impact of the QOF.

The majority of evaluations focusing on equity indicate that the QOF has encouraged a greater consistency of care across patient characteristics such as age, ethnicity and deprivation, but concerns remain over the use of exception reporting of patients (Gillam et al., 2012). Patient experience appears to have deteriorated as a result of the focus of QOF indicators on biomedical care, with some studies showing worsening of continuity of care and poorer access to preferred choice of practitioner. The policy does, however, appear to have improved practice organisation and resulted in a more appropriate use of skill-mix.

With regards to efficiency, Gillam et al. (2012) conclude that the evidence that improvements in care for chronic conditions in general practice have led to reductions in emergency hospital admissions is limited. Walker et al. (2010) considered the cost-effectiveness of the policy by comparing the size of the reward payments to the value of the expected health gains, based on published cost-effectiveness evidence. This was however only available for nine of the 146 original QOF indicators. They concluded that for this subset of indicators, the QOF incentive payments would be cost-effective even if they resulted in only modest improvements in performance. However, theirs was not a full cost-effectiveness analysis even within these indicators as they did not examine actual health gains and failed to take account of the administration costs of the scheme (Meacock et al., 2014).

A great deal of concern has been expressed about the potential for practices to 'game' the QOF because the performance data are extracted from their internal recording systems. Practices decide which patients appear in the denominator for each indicator as well as in the numerator through their diagnostic practice and their reporting of patients as 'exceptions'. In general, the view is that practices have not exploited this possibility *en masse*, and 'gaming' is not a general problem. Nevertheless, Gravelle et al. (2010) demonstrated that a minority of practices were behaving strategically to maximise their QOF payments by increasing prevalence if they were above the upper threshold and

increasing exceptions if they were below the upper threshold. In addition, Kontopantelis et al. (2012) found that a 5% increase in the upper payment threshold for one of the indicators in 2006/7 led to a 0.41% increase in the proportion of patients meeting the indicator and a 0.26% increase in the proportion of patients who were exception reported. The magnitude of the intended effect was larger than the gaming effect but the gaming effect was, nevertheless, non-negligible. It may be that practices have not exploited the full potential for gaming because the quality standards required were relatively easy to achieve, but that they will engage in more gaming if the performance requirements become more stringent.

3.2 Advancing Quality

There have been two evaluations of the AQ initiative to date, both of which take advantage of the fact that AQ was implemented in one region of England only, and achieved universal participation within this area, to employ a quasi-experimental design rarely available in this area of research. Sutton et al. (2012) employ a triple difference-in-differences methodology to examine the effects of AQ on mortality rates for three of the five conditions incentivised under the policy: AMI, heart failure, and pneumonia. They compare changes over time in outcomes of patients treated for the three incentivised conditions within the North West with changes in two other groups: patients treated for the same conditions in the rest of England, and patients treated for selected non-incentivised control conditions in the country as a whole. Their evaluation concludes that risk-adjusted mortality rates reduced significantly during the 18-month period under evaluation, with an absolute rate reduction of 1.3 percentage-points and a relative risk reduction of 6%. They estimate this to be equivalent to 890 fewer deaths amongst a patient population of 71,000 as a result of the policy.

The second evaluation of AQ builds upon this work to examine the often neglected topic of the cost-effectiveness of P4P policies. Meacock et al. (2014) employed a between-region difference-in-differences methodology comparing changes in the outcomes of patients treated for the same three incentivised conditions within the North West to changes in those of patients treated for these conditions in the rest of England. The outcomes studied are risk-adjusted mortality rates, emergency readmissions and length of stay (LOS). They find a significant reduction in mortality and LOS associated with the policy, whilst readmission rates remained unchanged. This mortality reduction was then converted into an estimate of the Discounted and Quality-Adjusted Life Expectancy (DANQALE) gains resulting from the policy. At a total programme cost of £13m, they conclude AQ to have represented a cost-effective use of resources from the commissioners' perspective during the 18-month period evaluated. They also estimate a cost saving to providers of £4.4m, since providers retain the cost savings from reducing length of stay under the existing tariff-based reimbursement system.

These results are at odds with those found for the HQID, which failed to demonstrate a significant effect on mortality within either its first three (Glickman et al., 2007; Ryan, 2009) or six years (Jha et al., 2012) of operation. The more positive results found in England as compared to the US have been attributed to the universal participation of providers within the region, the larger bonus payments and the collaborative nature of the scheme, which resulted in hospitals implementing wider general investments in quality improvement.

3.3 Commissioning for Quality and Innovation

The local design of CQUIN schemes makes it particularly difficult to evaluate the overall effects of this scheme due to the large variety of schemes and performance indicators, which complicates the identification of relevant comparison groups, and a considerable risk of selection bias in effect estimates. An evaluation of changes to health outcomes at the local level after CQUIN found no association between the inclusion of knee fracture as a CQUIN and patients' return to their usual place of residence after admission for such a fracture (McDonald et al., 2013). The same report found no association between the inclusion of discharge planning in the CQUIN targets and readmission rates for knee fracture. The evaluation also failed to detect any change in patient reported health gain after including groin, hernia, varicose vein, or hip and knee replacement surgeries as topics in local schemes. Finally, the addition of a patient safety or risk assessment goal was not associated with a change in the number of patient safety incidents recorded. The only positive association detected was between the inclusion of hip fracture as a focus area and the subsequent probability of return to a patient's usual place of residence. These results should, however, be taken with caution as no account was taken of the potential for selection bias.

With respect to the policy goals of generating enthusiasm among front line staff and identifying local needs and priorities for quality improvement, Kristensen et al. (2013) found the scheme to be largely unsuccessful. Whilst the programme had managed to identify local priorities, the desired enthusiasm did not manifest, and the resulting local CQUIN policies did not meet the design requirements for coverage and indicator choice set out by the Department of Health.

3.4 Best Practice Tariffs

McDonald et al (2012) were commissioned by the Department of Health to evaluate the impact of the BPTs introduced during the first year of the policy. For the increase in price for performing gall bladder removal as a daycase procedure, which applied nationally, they made comparisons to a basket of procedures which could also potentially be undertaken as daycases. They found that the daycase BPT led to a 7 percentage point increase in the daycase rate. There was no evidence that providers responded by selecting patients more amenable to daycase treatment, by reducing quality or by increasing volume. However, patients had to wait an additional 14 days for treatment on average, which was likely due to the increased demand on daycase facilities.

For stroke and hip fracture, McDonald and colleagues exploited the fact that only around half of providers could receive the BPT payment. They examined changes in process measures closely linked to the BPT and three outcome measures: mortality within 30 days, readmission within 30 days and return to usual place of residence within 56 days. Their difference-in-differences results indicated that the BPT for stroke had no impact on the process and outcome indicators in the first year. In contrast, the hip fracture BPT had substantial effects, being associated with: a 4 percentage point increase in receipt of surgery within 48 hours of admission, a 0.7 percentage point decrease in mortality, and a 2.1 percentage point increase in the proportion of patients discharged home within 56 days. The differences in impact between these two BPTs may be attributable to the different structures of the tariffs, as the BPT for hip fracture is only paid if all criteria are met whereas providers are rewarded separately for each indicator in the stroke BPT. It may also be attributable to differences in underlying trends in quality, concomitant service reconfigurations in stroke, and the emergence of a new drive to improve hip fracture care

focused on a new national audit programme which supplied the performance data for paying the hip fracture BPT.

3.5 Non-payment policies

As non-payment policies were only introduced in England in 2011, formal evidence of their impact on health outcomes has yet to be published. These programmes have, however, been met with resistance from the medical community (Diggory, 2010; Harrop-Griffiths, 2011; Lloyd and Maxwell-Armstrong, 2010). A prospective study (Kristensen et al., 2012) used a large administrative data set prior to the introduction of the policy to identify factors influencing readmissions that were both within and outside a hospital's control, and assessed the fraction of unexplained variance in readmission rates between providers. This unexplained variation was found to be very small, and patient characteristics observable at the time of admission found to be strong predictors of readmission probabilities. It was therefore hypothesised that non-payment policies could lead to access problems for patients displaying such observable characteristics.

Evidence on the impact of the English Never Events policy is even sparser. Recent publications from the English Department of Health report an increase in the number of Never Events from 166 in 2009/10 (Department of Health, 2012b) to 329 in 2012/13 (NHS England, 2013a)¹. The list of Never Events has however expanded over time, making such comparisons relatively uninformative.

Across the years, the majority of reported never events relate to retained foreign object post-procedure, and wrong site surgery.

4 Lessons

The wide range of P4P programmes introduced in England over the past decade illustrates the high level of uncertainty over how best to design such schemes. Systematic reviews of the effects of P4P offer little definitive guidance on this issue, instead highlighting the importance of context and the heterogeneity of the impacts found among the few schemes that have been evaluated robustly (Eijkenaar et al., 2013; Van Herck et al., 2010). There is also a lack of good theoretical models of P4P. It therefore seems that there is little information available to guide policy makers on how best to design P4P initiatives that will have predictable effects.

Nonetheless, the English experience can offer some general lessons. The first of these relates to the way in which P4P schemes should be introduced, and the influence which this has upon the ability to evaluate their effects. There has rarely been piloting of programmes, and schemes are often introduced simultaneously with new performance measurement, monitoring systems and new data infrastructures. Evaluations are therefore plagued by a lack both of comparable data before and after introduction and of control units not exposed to the initiative.

It does appear to be clear that the introduction of P4P in isolation is less effective. The most promising examples of English schemes are those that were introduced as part of wider quality improvement initiatives. For example, both AQ and the BPT for fragility hip fracture involved preparatory work in the form of structured clinical audits and learning collaboratives. Additionally, it seems that publishing providers' performance data adds an important reputational dimension to the incentive programmes.

¹ There is some uncertainty attached to the methods of counting never events, as Never Events in practice are reported through two information channels with some but not perfect overlap. The above number could thus be slightly higher.

The way in which P4P schemes have been described by health care professionals in qualitative work suggests that they are seen by clinicians not as financial incentives, but as an offer of supporting investment which can be used in negotiations with Directors of Finance when making the business case for investments in staff, equipment or systems (McDonald et al., 2013, 2012). This may oppose the standard ‘rational’ costs and income argument posed by economists, as the size of potential rewards are often smaller than the cost of the investment required to make the quality improvements necessary to receive the bonus payments. P4P may therefore be simply one of many mechanisms needed to facilitate quality improvement.

Further economically ‘irrational’ behaviours have been observed in response to P4P programmes in England. For example, although AQ began as a pure tournament system in which only the top 50% of participating providers could earn bonuses, representatives from all hospitals actively engaged in collaborative learning events in which successful hospitals shared lessons on how they achieved their high performance with all other teams in the region. Such collaboration ensured that the region made substantial performance improvements overall, but reduced the probability that the leading hospitals would remain at the top of the league table and therefore continue to earn financial rewards. It has also been demonstrated that in some instances, P4P schemes have led to quality-improving behaviours that have actually reduced providers’ own costs, and would have therefore likely been in their financial interests to implement even in the absence of the policy. Meacock et al. (2014) found that the quality improvements implemented by hospitals in response to AQ resulted in significant reductions to length of stay, the financial savings from which would have been accrued to providers rather than payers. These savings of £4.4m were similar in magnitude to the £4.8m available in rewards. McDonald et al. (2012) showed that the BPT for gall bladder removal led hospitals to increase their day case rate for this surgery, which was again expected to be cost-saving from the hospitals’ perspective.

England’s experience of P4P also suggests that fears over major unintended consequences may be unjustified in this sector. While general practices had strong incentives to engage in gaming of the QOF, the empirical evidence suggests that this happened only to a small extent and in a minority of providers (Doran et al., 2008; Gravelle et al., 2010). Evaluations of other P4P schemes also reveal little evidence of unintended consequences. A particular concern had been the equity impact, with a fear that providers would engage in cherry-picking or may prioritise patients for which the quality metrics were easier or less costly to achieve. Studies examining equity with regards to the QOF, summarised by Gillam et al. (2012), instead find that care became more consistent across patient groups. The P4P programmes introduced in England may have had limited scope or may not have been sufficiently challenging to produce such behaviours or the intrinsic, altruistic motives of public health care providers may prevent such negative reactions. Hospitals’ lack of responses to the cataract surgery BPT and practices’ low rate of uptake of the depression QOF indicators support this explanation, indicating that providers will not respond to P4P if the required activities clash with their own values. This suggests that incentives must be aligned with healthcare professionals’ general motivation and beliefs if they are to be effective.

A further encouraging message from the English experience is that P4P programmes may represent a cost-effective use of resources compared to existing FFS or capitation schemes. Modelling exercises for the QOF indicators, for example, showed that small changes in practice performance on the incentivised metrics for which there was good quality cost-effectiveness evidence available could make the overall scheme cost-effective (Walker et al., 2010). The potential cost-savings resulting from avoided hospital

admissions and the QALY gains from secondary prevention activity in the community were estimated to exceed what were large relative increases in income for general practices, but small amounts in terms of the global budget. Meacock et al.'s (2014) analysis of AQ drew similar conclusions, demonstrating that even small mortality reductions achieved across a large patient population can significantly outweigh the costs of implementing a P4P programme from the commissioner's perspective.

There are still, nonetheless, many gaps in the P4P literature. Little work has focused on how to systematically set the price levels in financial incentive schemes. Kristensen et al. (2013) suggest that information on several important aspects is required, such as the level of provider altruism, providers' costs, purchasers' willingness to pay, and the opportunity cost of public funds. These are all critical to an effective price-setting formula, yet the information requirements to obtain such estimates are daunting. There is also an absence of head-to-head comparisons of the cost-effectiveness of P4P compared to other ways to improve quality, such as targets, public reporting and multi-faceted quality improvement programmes.

Finally, the long-term prospects for P4P programmes are unknown. The literature suggests that the effects of such policies are modest and short-term. The potential gains fall over time as providers approach a natural maximum level of achievable performance, yet the size of bonus payments is often kept constant. Purchasers do not currently appear to have an exit strategy once P4P has been introduced, and evidence on the effect of withdrawing payments is sparse. Analysis by Lester et al. (2010) suggests that the subsequent performance declines may be very large, and even fall below levels observed before the introduction of the policy. This is at odds however with the general perception that quality improvement is an investment rather than a transitory activity.

5 The future of pay-for-performance in England

Some indications of the direction of future policy are the emerging trends in P4P programmes introduced to date. Table 1 shows how the key features of the P4P schemes in England have evolved over time. The size of the financial incentives appears to be increasing, and there has been a movement from bonuses to penalties. Rewards are most commonly based upon absolute performance, despite the most positive findings coming from the AQ initiative, which rewarded providers according to their relative performance. There is evidence also of a growing use of administrative data rather than self-reported for the monitoring of schemes, with the frequency of this monitoring increasing. Most recent policies in England are being implemented through mandatory participation, with a distinct lack of supporting levers.

Other indications of the direction of future policy can be seen in policy documents, which suggest that the use of P4P is set to further increase across England. A recent discussion paper published by the new national body for the NHS in England (*NHS England*) states that financial incentives "should underpin the delivery of NHS England's strategy", with the aims of improving patient outcomes, reducing health inequalities, and ensuring the maintenance of basic quality standards (NHS England, 2013b). This report was published as part of a wider review of pay-for-performance schemes, and is being used to inform the planning round for NHS contracts in 2014/15. NHS England have indicated that the long term direction of travel on incentives is towards providers receiving a core payment for a specified quantity of service provision, with the additional opportunity to earn a substantial further percentage reward for meeting various standards (NHS England, 2013b). It has been recognised that financial incentives cannot deliver these aims alone, and so will continue to be used in partnership with measures such as

sharing best practice and performance benchmarking. There are limits to the potential coverage of P4P because such schemes require that quality is known and measurable and able to be monitored at a reasonable cost. This is especially germane to hospital care, where it is notable that P4P schemes have focused on a relatively narrow range of health conditions and interventions. A move is also being made towards a more cohesive national approach, after problems arising from the locally negotiated CQUIN agreements.

The latest annual negotiation of the national GP contract has resulted in a substantial reduction in the amount of income linked to performance. The funding will instead be spent on paying practices to provide enhanced services or to increase capitation payments. The rationale provided is that this will reduce bureaucracy and allow care to be focused on the needs of individual patients, perhaps signalling a loss of enthusiasm for P4P in primary care. Enthusiasm does, however, still appear to be present in the secondary care sector. The national CQUIN policy is set to continue, with the value of the initiative remaining constant at 2.5% of the national contract value for 2014/15. The design of local schemes is set to change though, with a more rigorous approach being applied to the chosen indicators. It has been suggested that this will either be achieved by a reduction in the number of indicators chosen locally, publication of clearer rules around indicator development, or the introduction of a pick-list from which local providers and commissioners can select nationally-approved quality metrics upon which to base financial rewards. Under the umbrella of CQUIN, hospitals in the North West region of England are continuing to use the original AQ quality metrics, and expanding this initiative cover whole care pathways and additional clinical areas (Advancing Quality Alliance, 2013).

The use of financial incentives is also set to expand into commissioners' contracts. Clinical Commissioning Groups (CCGs), the new membership organisations of general practices responsible for planning and purchasing services for local populations, are set to have a 'Quality Premium' introduced into their annual budgets. This premium will reward CCGs for improvements in the quality of services which they commission, and the gains in health outcomes and inequalities associated with these improvements (NHS England, 2013c). Local Authorities, who have new responsibilities for improving local public health, will also have a 'Health Premium' from 2014/15.

As the current P4P schemes continue across England more evidence will emerge on the long-term effects of these schemes, as well as the most effective way to design such programmes. As the Nordic countries increase their use of P4P programmes they can also contribute to the body of evidence, if national programmes are introduced in a way that facilitates evaluation.

References

- Advancing Quality Alliance, 2013. Advancing Quality 3 Year Business Plan 2013/14 - 2015/16.
- Anell, A., Glenngård, A., Merkur, S., 2012. Sweden: Health system review., Health Systems in Transition.
- Department of Health, 2008a. Using the Commissioning for Quality and Innovation (CQUIN) Payment Framework. London.
- Department of Health, 2008b. High quality care for all: NHS next stage review final report. London.
- Department of Health, 2008c. Impact Assessment of Commissioning for Quality and Innovation payment framework.
- Department of Health, 2011. Payment by Results Guidance for 2011-12. Leeds.
- Department of Health, 2012a. The “Never Events” list 2012/13. London.
- Department of Health, 2012b. The Never Events Policy Framework: an update to the never events policy. London.
- Department of Health, 2012c. Payment by Results Guidance for 2012-13. Leeds.
- Department of Health, 2013. Payment by Results Guidance for 2013-14. Leeds.
- Diggory, P., 2010. Re:Hospitals will be fined for emergency readmissions, says Lansley [WWW Document]. BMJ Br. Med. J. - Rapid Responses. URL <http://www.bmj.com/rapid-response/2011/11/02/readmission-hospital-not-always-failure-care>
- Doran, T., Fullwood, C., Reeves, D., Gravelle, H., Roland, M., 2008. Exclusion of Patients from Pay-for-Performance Targets by English Physicians. N. Engl. J. Med. 359, 274–284. doi:10.1056/NEJMsa0800310
- Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O., 2013. Effects of pay for performance in health care: A systematic review of systematic reviews. Health Policy 110, 115–130. doi:10.1016/j.healthpol.2013.01.008
- Gillam, S.J., Siriwardena, A.N., Steel, N., 2012. Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework—A Systematic Review. Ann. Fam. Med. 10, 461–468. doi:10.1370/afm.1377
- Glickman, S.W., Ou, F.-S., DeLong, E.R., Roe, M.T., Lytle, B.L., Mulgund, J., Rumsfeld, J.S., Gibler, W.B., Ohman, E.M., Schulman, K.A., Peterson, E.D., 2007. Pay for performance, quality of care, and outcomes in acute myocardial infarction. JAMA J. Am. Med. Assoc. 297, 2373–2380. doi:10.1001/jama.297.21.2373
- Gravelle, H., Sutton, M., Ma, A., 2010. Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?*. Econ. J. 120, F129–F156. doi:10.1111/j.1468-0297.2009.02340.x
- Greene, S.E., Nash, D.B., 2009. Pay for Performance: An Overview of the Literature. Am. J. Med. Qual. 24, 140–163. doi:10.1177/1062860608326517
- Harrop-Griffiths, W., 2011. Never Events. Anaesthesia 66, 158–162. doi:10.1111/j.1365-2044.2010.06624.x
- Helsedirektoratet, 2013a. En gjennomgang av kvalitetsbaserte finansieringsordninger [A review of quality based reimbursement schemes] (No. IS-1940). Oslo.

- Helsedirektoratet, 2013b. Kvalitetsbasert finansiering 2014 [Quality based reimbursement 2014] (No. IS-2111). Oslo.
- Jha, A.K., Joynt, K.E., Orav, E.J., Epstein, A.M., 2012. The Long-Term Effect of Premier Pay for Performance on Patient Outcomes. *N. Engl. J. Med.* 366, 1606–1615. doi:10.1056/NEJMSa1112351
- Kontopantelis, E., Doran, T., Gravelle, H., Goudie, R., Siciliani, L., Sutton, M., 2012. Family Doctor Responses to Changes in Incentives for Influenza Immunization under the U.K. Quality and Outcomes Framework Pay-for-Performance Scheme. *Health Serv. Res.* 47, 1117–1136. doi:10.1111/j.1475-6773.2011.01362.x
- Kristensen, S.R., McDonald, R., Sutton, M., 2013. Should pay-for-performance schemes be locally designed? evidence from the commissioning for quality and innovation (CQUIN) framework. *J. Health Serv. Res. Policy* 18, 38–49. doi:10.1177/1355819613490148
- Kristensen, S.R., Sutton, M., Bech, M., 2012. Holding hospitals financially responsible for emergency readmission rates - potential for access problems?, in: Kristensen, S.R. (Ed.), *Performance Based Provider Reimbursement and Supply Side Incentives for Quality in Hospital Care*. PhD Dissertation. University of Southern Denmark, Odense.
- Lester, H., Schmittiel, J., Selby, J., Fireman, B., Campbell, S., Lee, J., Whippy, A., Madvig, P., 2010. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ* 340, c1898–c1898. doi:10.1136/bmj.c1898
- Lloyd, G.M., Maxwell-Armstrong, C., 2010. Re:Hospitals will be fined for emergency readmissions, says Lansley [WWW Document]. *BMJ Br. Med. J. - Rapid Responses*. URL <http://www.bmj.com/rapid-response/2011/11/03/rehospitals-will-be-fined-emergency-readmissions-says-lansley>
- Maynard, A., 2012. The powers and pitfalls of payment for performance. *Health Econ.* 21, 3–12. doi:10.1002/hec.1810
- McDonald, R., Kristensen, S.R., Zaidi, S., Sutton, M., Todd, S., Konteh, F., Hussein, K., Brown, S., 2013. Evaluation of the Commissioning for Quality and Innovation Framework Final Report.
- McDonald, R., Zaidi, S., Todd, S., Konteh, F., Hussain, K., Roe, J., Allen, T., Fichera, E., Sutton, M., 2012. A Qualitative and Quantitative Evaluation of the Introduction of Best Practice Tariffs.
- Meacock, R., Kristensen, S.R., Sutton, M., 2014. The Cost-Effectiveness of Using Financial Incentives to Improve Provider Quality: A Framework and Application. *Health Econ.* 23, 1–13. doi:10.1002/hec.2978
- Ministeriet for Sundhed og Forebyggelse, 2013. Bedre Incitamentter i Sundhedsvæsenet [Better Incentives in the national health service]. Copenhagen.
- NHS England, 2013a. Never events data summary for 2012/2013. London.
- NHS England, 2013b. Review of incentives, rewards and sanctions: Discussion paper for stakeholders.
- NHS England, 2013c. Quality Premium: 2013/14 guidance for CCGs.
- NPSA, 2009. Never Events Framework 2009/10. National Reporting and Learning Service - National Patient Safety Agency.

- Roland, M., 2004. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. *N. Engl. J. Med.* 351, 1448–1454. doi:10.1056/NEJMhpr041294
- Rosenthal, M.B., 2007. Nonpayment for performance? Medicare's new reimbursement rule. *N. Engl. J. Med.* 357, 1573.
- Ryan, A.M., 2009. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost. *Health Serv. Res.* 44, 821–842. doi:10.1111/j.1475-6773.2009.00956.x
- Sutton, M., Elder, R., Guthrie, B., Watt, G., 2010. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Econ.* 19, 1–13. doi:10.1002/hecl.1440
- Sutton, M., Nikolova, S., Boaden, R., Lester, H., McDonald, R., Roland, M., 2012. Reduced Mortality with Hospital Pay for Performance in England. *N. Engl. J. Med.* 367, 1821–1828. doi:10.1056/NEJMsa1114951
- Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M.B., Sermeus, W., 2010. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv. Res.* 10, 247. doi:10.1186/1472-6963-10-247
- Walker, S., Mason, A.R., Claxton, K., Cookson, R., Fenwick, E., Fleetcroft, R., Sculpher, M., 2010. Value for money and the Quality and Outcomes Framework in primary care in the UK NHS. *Br. J. Gen. Pract.* 60, e213–e220. doi:10.3399/bjgp10X501859