HEALTH ECONOMICS
Health Econ. **23**: 1–13 (2014)
Published online 14 August 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.2978

THE COST-EFFECTIVENESS OF USING FINANCIAL INCENTIVES TO IMPROVE PROVIDER QUALITY: A FRAMEWORK AND APPLICATION

RACHEL MEACOCK[a,*], SØREN RUD KRISTENSEN[a,b] and MATT SUTTON[a]

[a]*Manchester Centre for Health Economics, Institute of Population Health, The University of Manchester, UK*
[b]*COHERE, Department of Business and Economics, University of Southern Denmark, Denmark*

ABSTRACT

Despite growing adoption of pay-for-performance (P4P) programmes in health care, there is remarkably little evidence on the cost-effectiveness of such schemes. We review the limited number of previous studies and critique the frameworks adopted and the narrow range of costs and outcomes considered, before proposing a new more comprehensive framework, which we apply to the first P4P scheme introduced for hospitals in England. We emphasise that evaluations of cost-effectiveness need to consider who the residual claimant is on any cost savings, the possibility of positive and negative spillovers, and whether performance improvement is a transitory or investment activity. Our application to the *Advancing Quality* initiative demonstrates that the incentive payments represented less than half of the £13m total programme costs. By generating approximately 5200 quality-adjusted life years and £4.4m of savings in reduced length of stay, we find that the programme was a cost-effective use of resources in its first 18 months. Copyright © 2013 John Wiley & Sons, Ltd.

Received 22 February 2013; Revised 16 May 2013; Accepted 02 July 2013

KEY WORDS: Pay-for-performance; Cost-effectiveness

1. INTRODUCTION

Pay-for-performance (P4P) schemes, which link financial payments by purchasers to the quality of care supplied by health care providers, have grown in popularity over recent years. Care quality is commonly measured using pre-specified performance measures, which are often clinical processes judged to represent best practice or, less frequently, measures of outcome. Where clinical process measures are used, it is hoped that this will produce superior health outcomes for patients. Improving quality and outcomes may also reduce future health care costs. Despite much research by economists on this topic, there remains remarkably little evidence on the cost-effectiveness of such schemes.

A recent commentary (Maynard, 2012) highlights the 'curious' focus of research to date on the effectiveness of P4P schemes, with a neglect of their costs and therefore cost-effectiveness. This gap in the evidence base is also noted by a number of reviews. Greene and Nash (2009) provided an overview of the literature on P4P published between 2004 and 2008. Of the 100 articles included in their annotated bibliography, only three are grouped under the heading of 'cost analysis' (Curtin *et al*., 2006; Nahra *et al*., 2006; Parke, 2007). Mehrotra and colleagues (2009) systematically reviewed the evidence on hospital-based P4P programmes, stating there to be approximately 40 of such schemes targeted at inpatient care. Despite this, only eight formal evaluations were found, covering just three different schemes. Of these eight published studies, just one (Nahra *et al*., 2006) attempted to estimate cost-effectiveness.

*Correspondence to: Manchester Centre for Health Economics, Institute of Population Health, The University of Manchester, 4.306 Jean McFarlane Building, Oxford Road, Manchester, M13 9PL, UK. E-mail: Rachel.meacock@manchester.ac.uk

Most recently, Emmert *et al*. (2012) presented a systematic review of economic evaluations of P4P, critically assessing the identified studies on their methodological quality according to the widely used Drummond and Jefferson (1996) checklist. They identify nine studies, of which only three were categorised as full economic evaluations[1] and six as partial evaluations.[2] Of these six, four were deemed to be partial evaluations as they examined both the costs and effects of the P4P programmes under consideration but failed to make an explicit link between the two.[3] The remaining two partial evaluations were simple cost comparisons, examining only the financial implications of the schemes in question.[4] This comprehensive review concluded that, on the whole, studies to date are methodologically flawed, failing to incorporate the full range of costs and consequences relevant to the evaluation of P4P.

Concerns regarding the value for money of the Quality and Outcomes Framework (QOF) in the UK led the Department of Health to commission a report in which a conceptual framework was developed to assess the cost-effectiveness of QOF indicators (Mason *et al*., 2008; Walker *et al*., 2010). This framework takes account of the cost of providing the incentivised interventions along with the incentive payments and the value of the health benefits achieved but fails to incorporate the administrative costs associated with running the scheme. It also only considers the direct costs and benefits of changes in the incentivised measures and does not account for other changes in provider behaviour. Finally, it simulates the effects of better performance on the incentivised measures using published estimates of average effects and therefore does not reflect incremental changes. Although it is fundamentally important to ensure that the treatments incentivised by P4P programmes are themselves cost-effective, even after the additional cost of the incentive payments are considered, we believe it is necessary to take this a step further and consider whether P4P programmes as a whole represent a cost-effective use of resources.

Therefore, we aim to develop an analytical framework to guide the assessment of the cost-effectiveness of P4P programmes, highlighting the issues that should be considered when undertaking such evaluations. We first critique the narrow range of costs and effects considered by studies to date. This leads us to propose a new more comprehensive framework, highlighting the various cost categories that should be considered beyond the incentive payments themselves, along with issues such as who the residual claimant on any cost savings may be. Finally, we apply this framework to the first P4P scheme introduced for hospitals in the UK, the Advancing Quality (AQ) programme. The introduction of this scheme has been shown to have been associated with a significant reduction in mortality in the short term (Sutton *et al*., 2012). We use our framework to show what additional analyses are required to assess whether the scheme was cost-effective. In particular, we consider how to convert the mortality reductions to gains in quality-adjusted life years (QALYs), what direct set-up and running costs to include and estimate other indirect impacts on health service costs.

2. METHODS

2.1. Critiquing previous evaluations

The recently published Emmert *et al*. (2012) review systematically appraised the quality of the economic evaluation literature. We present a brief commentary on the lack of methodological consistency between studies, focusing on the narrow range of costs and effects considered. Studies were identified from the previously mentioned review, and the search strategy used was run again in September 2012 to ensure that no new articles were missed. Studies known to the authors but not included in the Emmert *et al*. (2012) review were also included if they assessed the costs of P4P schemes. Details were extracted regarding the setting of the evaluation, the perspective taken, the main cost categories included and omitted, and the outcomes examined.

[1]Kouides *et al*., 1998; Nahra *et al*., 2006; An *et al*., 2008.
[2]Norton, 1992; Curtin *et al*., 2006; Parke, 2007; Rosenthal *et al*., 2009; Ryan, 2009; Lee *et al*., 2010.
[3]Norton, 1992; Rosenthal *et al*., 2009; Ryan, 2009; Lee *et al*., 2010.
[4]Curtin *et al*., 2006; Parke, 2007.

DOI: 10.1002/hec

2.2. Developing the analytical framework

This appraisal of previously published evaluations was then used to develop a more comprehensive framework for assessing the cost-effectiveness of P4P schemes. The methodological issues brought to light in this first section were combined with the standard principles of cost-effectiveness analysis outlined in already established frameworks (e.g. Drummond *et al*., 2005; NICE, 2013) to provide a more specific framework to guide the evaluation of P4P programmes.

2.3. Applying the framework to the Advancing Quality initiative

We demonstrate our proposed framework by applying it to the AQ initiative that began in the North West of England in October 2008. We focus on the first 18 months, after which it was absorbed into the national Commissioning for Quality and Innovation scheme (Department of Health, 2008). The programme aimed to improve quality in participating hospitals by paying for performance on 28 indicators across five health conditions. AQ ran in the North West of England only, and participation was universal within this region. We first discuss the issues raised in our framework in relation to the evaluation of AQ, before presenting estimates of the cost and effects of the programme.

We analyse mortality within 30 days of admission, emergency readmissions within 30 days and length of stay (LOS) for three of the five incentivised conditions (acute myocardial infarction (AMI), heart failure and pneumonia). We exclude coronary artery bypass grafting (CABG) and hip and knee replacement as the mortality rate was below 2% for these procedures during the pre-intervention period. We use patient level Hospital Episode Statistics data for patients admitted for one of these three AQ conditions in the period 1st of April 2007 to 31st of March 2010, covering 18 months before and 18 months after the introduction of the programme. For the analysis of readmissions, we also include readmissions that occurred in April 2010. Our sample consists of 856 715 patients (662 458 patients for readmissions as we exclude patients not discharged alive) treated for one of the three conditions we examine at one of 154 hospitals across England. Of these, 24 hospitals were in the North West of England, and thus subject to AQ, with the remaining 130 located in other regions of England, and therefore not subject to the policy. We evaluate the effects of AQ using a between-region difference-in-differences (DID) analysis, comparing changes in outcomes in the North West to the changes in outcomes in the rest of England. The analysis was carried out at hospital level using weighted least squares on quarterly observations of risk-adjusted in-hospital mortality, readmission and mean LOS, allowing for hospital fixed effects and for time trends using quarterly dummy variables. The risk adjustment for each of the three outcomes of interest was conducted at patient level. Our model for identifying changes in outcome after the introduction of AQ takes the form:

$$y_{jt} = a + u_j + v_t + \delta D_t^1 \times D_t^2 + \varepsilon_{jt}$$

with y_{jt} being the risk-adjusted outcome of interest at hospital j in quarter t, u_j the hospital fixed effects, v_t the time fixed effects and ε_{jt} the residual term that is randomly distributed with a zero mean. The dummy variable D_j^1 equals 1 if the hospital is located in the North West and zero otherwise. The variable D_t^2 equals 1 for all quarters after the introduction of AQ and zero beforehand. Our main interest is in the coefficient on the interaction of these two variables, δ. The main effects of D_j^1 and D_t^2 are not included, as they are perfectly collinear with the included time and hospital fixed effects.

As well as considering changes in these variables in natural units, we repeat the DID estimation using variables to which 'tariffs' have been applied. We apply a discounted and quality-adjusted life expectancy (DANQALE) tariff to the mortality outcome and the cost tariffs used in the national activity-based financing programme ('Payment by Results') to the readmissions and LOS.

The DANQALE tariff is stratified by single year of age (18–100 years) and sex. Sex-specific life expectancy estimates at each single year of age are taken from the 2008–2010 Interim Life Tables from the ONS (2011). The age-sex specific quality of life adjustments are sourced from mean values of the EQ-5D index reported by

DOI: 10.1002/hec

respondents to the 2006 wave of the Health Survey for England. We calculate the DANQALE (Q_{ia}) for each individual i in each age-sex group a as:

$$Q_{ia} = (1 - m_i)\sum_{j=a}^{L_a} q_j(1 + r)^{-(j-a)}$$

where m_i equals 1 if the individual dies within 30 days and 0 otherwise; j indexes ages from age a to the life expectancy of an individual currently aged a (L_a); q_j is health-related quality of life at age j; and r is the discount rate. We use an annual discount rate of 3.5% as specified by the National Institute for Health and Care Excellence (NICE) in their reference case (NICE, 2013). To cost LOS, we apply to each individual's LOS the 2009/10 per diem tariffs for days above the trim point for the main healthcare resource group (HRG) for which they were admitted. Readmissions are costed using the 2009/10 tariff prices for the main HRG for which the individual is admitted on readmission.

A critical assumption of DID is that the changes in the control hospitals are an appropriate counterfactual for the changes in the treated hospitals that would have occurred without the programme. We undertook pre-trends test for all of the raw and tariffed outcomes and failed to reject the null-hypothesis of equal pre-trends at the 5%-significance level for all conditions and outcomes bar LOS for heart failure patients (Table I).

3. RESULTS

3.1. Previous evaluations

We identified 14 studies examining the cost of P4P schemes (Table A1). The majority of these schemes operated in the USA [1–8 and 10–11],[5] with two in the UK [13–14] and one in each of Germany [9] and China [12]. The most common setting for the programmes under evaluation were primary care clinics [2, 4, 6–9 and 12–14], followed by hospitals [3, 5 and 11]. Nine of the 10 US evaluations were undertaken from the perspective of the health plan [1–5, 7–8 and 10–11], with one extending this to include the plan's enrolees [7] and another also considering the providers' perspective [1]. Just one evaluation was performed purely from the providers' perspective [9], and the remaining three from that of government-run health systems [12–14]. The range of costs included by many of the studies were however inconsistent with their stated perspectives, often failing to encompass all relevant cost categories. Just two evaluations clearly incorporated the costs associated with the development and set-up of the P4P schemes in question [4, 8], and only six included the ongoing running costs [4–9]. Seven studies made some attempt to measure the increased costs associated with providing the incentivised treatments [1, 4, 6, 8–9, 12 and 14]. Five studies failed to consider any costs beyond the incentive payments themselves [2–3, 10–11 and 13].

Of the 14 studies examining the cost of P4P programmes, 11 also made some attempt to estimate the effects of the schemes [1–3, 5 and 8–14]. However, the range of effects considered was narrow. The incentivised performance measures were by far the most commonly used metrics of effect, with all but one evaluation reporting results on these process or clinical measures [11]. Four evaluations considered only these incentivised measures [2–3 and 8–9] and made no attempt to link quality improvements to health outcomes. Three studies examined intermediate outcomes, such as hospitalisations and LOS [1, 10 and 12], and three examined mortality [1, 5 and 11]. Just two of the evaluations attempted to express the effects of P4P schemes in terms of QALYs [5 and 14], and only one looked at the potential effects on non-incentivised areas of care [13].

The omission of relevant cost categories by many previously conducted evaluations, along with the lack of evidence regarding the effects on health outcomes, means that conclusions regarding the value for money of the programmes in question cannot be made.

[5]Numbers in [] refer to the study number given in Table A1 and are used to enable ease of reading for this summary

DOI: 10.1002/hec

Table I. Descriptive statistics

Condition	North West region			Rest of England			Pre-trend tests	
	Before introduction	After introduction	Difference	Before introduction	After introduction	Difference		
AMI								
Patients	20 092	18 762	−1330	104 912	101 479	−3433		
Mortality rate	12.4	11.0	−1.4	11.0	10.7	−0.3	−0.4	[−1.02, 0.20]
Readmission rate	11.9	12.1	0.2	10.9	11.1	0.2	−0.3	[−0.91, 0.25]
Average LOS	9.3	8.5	−0.8	8.0	7.7	−0.3	−0.07	[−0.28, 0.14]
Heart failure								
Patients	15 446	15 476	30	83 546	86 569	3023		
Mortality rate	17.9	16.6	−1.3	16.6	16.1	−0.6	0.3	[−0.44, 1.02]
Readmission rate	17.8	18.4	0.7	17.3	17.0	−0.2	0.0009	[−0.80, 0.81]
Average LOS	11.9	11.2	−0.7	11.4	11.0	−0.5	−0.3	[−0.66, −0.04]
Pneumonia								
Patients	28 275	36 428	8153	150 526	195 204	44 678		
Mortality rate	28.0	25.9	−2.2	27.2	26.3	−0.9	−0.1	[−0.72, 0.46]
Readmission rate	15.1	15.7	0.6	13.2	13.7	0.5	−0.2	[−0.81, 0.42]
Average LOS	12.8	11.8	−1.0	11.8	11.4	−0.4	−0.2	[−0.47, 0.02]

LOS, length of stay.
The pre-trend tests are the estimated difference between the linear quarterly trends in the North West and rest of England. *95%* confidence intervals in brackets.

DOI: 10.1002/hec

3.2. Analytical framework

3.2.1. Perspective. The relevant perspective will depend upon the institutional arrangements into which the P4P programme is introduced. In the UK, this is likely to be that of the National Health Service (NHS) and personal social services, consistent with the perspective specified by NICE in their reference case (NICE, 2013). The perspective should be clearly stated, and the range of costs and effects considered should be consistent with this.

As health care providers act as agents to payers (who can be thought of as principals under Principal-Agent Theory), and these payers in turn act as agents to customers/taxpayers, it is worth at least considering the perspective of providers as well as payers when determining the cost-effectiveness of P4P. It may therefore be relevant to consider not only whether it is cost-effective for the payer to run a P4P programme but also whether it is cost-effective for providers to participate in and perform the tasks necessary to improve performance on the stipulated quality measures. This could be thought of as a business case perspective. Providers may incur substantial costs as a result of participating in P4P programmes, both in terms of the capital investments necessary to permit activities such as data collection and the cost of providing the incentivised treatment itself. Although some/all of these costs may be offset by the incentive payments, there is no guarantee that providers will actually receive bonuses as these are conditional upon performance. In some cases, such schemes operate as a 'tournament', with only the top performers receiving a bonus payment, and under some programmes, there may even be the possibility of financial sanctions if performance benchmarks are not met.

3.2.2. Comparator. A clear comparator is essential for any economic evaluation (Drummond *et al*., 2005), representing what would have happened in the absence of the programme. The relevant counterfactual will again depend upon the institutional arrangements. An important consideration is whether to compare with the same additional resources but paid in a different manner or whether to compare with no additional payments. This depends on whether we are interested in P4P as a payment mechanism or as a form of potential additional funding.

Ideally, the programme would first be introduced under conditions of randomisation, with providers being allocated to an intervention group receiving P4P or a control group. This would allow selection bias and confounding factors to be avoided. In practice, however, P4P schemes are rarely launched in this way (Scott *et al*., 2011). It may be possible to employ a quasi-experimental design using providers not participating in the scheme as a comparator group if, for example, P4P has only been implemented in certain geographical areas. It is vital, however, that the analysis takes into account any potential sources of bias such as differing provider or patient characteristics between the groups. Alternatively, providers may be used as their own controls in a before-after study design, with observed outcomes before the implementation of P4P being projected forward in order to predict outcomes in the absence of the programme. Again, attempts must be made to control for potential sources of bias such as general time trends, which may have also affected the outcomes under examination.

3.2.3. Cost categories. Although the incentive payments themselves are by far the most obvious cost component of P4P programmes, there are many other costs involved in design and implementation. Although their relevance and magnitude will differ between programmes, the following cost categories should be considered:

- Set up/development costs—e.g. staff time, infrastructure investment. These costs can be spread across the expected lifetime of the policy if this is known.
- Running costs—e.g. administration.
- Incentive payments
- Costs to providers of participating in the scheme—e.g. staff time, pharmaceuticals. The perspective of the evaluation will dictate whether these costs are relevant.
- Cost savings—e.g. reduced complications, LOS, readmissions. It is assumed that improving the quality of care will produce superior health outcomes, which in turn has consequences for future health care

DOI: 10.1002/hec

costs. These cost savings may fall on providers or commissioners depending on the payment rules, so it is important to consider who the residual claimant may be.

The previously mentioned cost categories and examples are not exhaustive and illustrate the many possible financial implications of P4P schemes beyond the incentive payments themselves. As with any economic evaluation, the likely magnitude of each cost category must be weighed up against the resources involved in accurate estimation. There may be justification for excluding certain costs if it is clear that either they are insignificant in comparison to the overall cost of the policy or their inclusion will simply further confirm the current conclusions, but this should nevertheless be discussed.

3.2.4. Opportunity cost. As with any economic evaluation, we are concerned with the opportunity cost of the resources used by a programme, which in the case of health care spending represents the possible health gains foregone through not providing alternative treatments. P4P programmes are not always financed by additional funds but may instead involve a reallocation of current budgets or resources. For example, a percentage of the existing budget may be top-sliced to fund the incentive scheme, or the duties of existing members of staff may be changed to focus on the areas of care incentivised. Although this does not involve any additional spending, these resources still have an opportunity cost in terms of care displaced.

3.2.5. Outcomes. The main outcomes recorded for P4P programmes are the targeted quality measures upon which performance is judged. If these are process rather than outcome measures, then evidence on their link with health outcomes should be presented. Ideally, benefits would be expressed in terms of QALYs in order to permit comparison with standard cost-effectiveness thresholds (NICE, 2013; Walker *et al*., 2010).

However, because quality is multi-dimensional, the outcomes influenced by P4P programmes are likely to stretch beyond those captured by the targeted performance measures, with the potential for both positive and negative spillovers into non-incentivised areas of care. If incentives divert the existing efforts of providers away from non-incentivised areas of care rather than promoting additional effort in the targeted areas, this could result in unintended consequences for patients (Kelman and Friedman, 2009). Depending on how well the chosen performance indicators capture the desired outcomes, the hospital's degree of altruism and to what extent effort on the incentivised and non-incentivised dimensions are substitutes or complements to the agent, it may even be undesirable to pay-for-performance (Holmstrom and Milgrom, 1991; Kaarboe and Siciliani, 2011). Gaming is also a possibility, where providers merely make their performance on the incentivised measures appear better than it actually is, normally through manipulation of the reporting systems used to record such performance. A broad range of outcomes extending beyond the incentivised measures should therefore be considered when evaluating P4P schemes in order to fully capture their effects, both intended and unintended.

3.2.6. Time horizon. As with any economic evaluation it is important to capture all of the relevant costs and consequences attributable to a programme, which are likely to span over a number of years. An interesting point to consider is the expected lifetime of P4P schemes, which are seldom stated, and their ability to induce continued quality improvements year on year. Although we may expect to observe performance improvements when P4P is first introduced, these may reach a ceiling after which little or no further quality improvements are achieved. It may then be relevant to consider the consequences of removing the financial incentives currently in place if they are failing to induce additional benefits. The effect of this removal will depend upon whether quality improvement is a transitory or investment activity. Quality could fall, perhaps even to levels below those observed before the introduction of P4P (Lester *et al*., 2010). Alternatively, incentivised behaviours may have become routine and therefore continue even after payments are withdrawn. If some of the benefits are sustained beyond the period of cessation of the incentive payments, the cost-effectiveness of the scheme will be underestimated by a restricted evaluation period.

DOI: 10.1002/hec

3.3. Application to Advancing Quality

3.3.1. Perspective. We examine the cost-effectiveness of AQ from the perspective of the NHS, estimating the costs incurred by commissioners and the resulting health benefits achieved. However, we note that, as the programme ran under a tournament system, only half of the hospitals received bonus payments at each payout. Providers may have incurred substantial participation costs yet received no financial rewards for their efforts.

3.3.2. Comparator. We take advantage of the fact that AQ was introduced through universal participation and in the North West of England only to employ a quasi-experimental design in which the rest of England acts as the comparator.

3.3.3. Cost categories. We seek to include all of the relevant costs incurred by commissioners. These include the one-off lump-sum grants given to providers to cover the investments in infrastructure necessary to enable the required data collection. As AQ was merged with another national P4P policy 18 months after its introduction, and the expected life time of this new policy is unknown, the entire set-up costs are attributed to the first 18 months. Thus, our estimate of the costs of AQ over the first 18 months represents the upper bound of the actual costs applicable to this period.

We also include the financial incentives paid out to providers, the ongoing running costs and other one-off costs incurred within the period. The general running costs include the contract with Premier Inc. who oversaw the scheme, the central AQ team, auditing activities, quality improving consultancies and other administrative costs. One-off costs include legal fees and other procurements. We examine the potential cost savings resulting from reduced readmissions and LOS and discuss who the residual claimants on these savings will have been.

3.3.4. Opportunity cost. AQ was financed by a reallocation of the North West commissioning budget and so did not result in any additional spending by payers. We cannot determine what this money would have been spent on in the absence of the policy and so use the standard UK cost-effectiveness threshold to reflect opportunity costs.

3.3.5. Outcomes. Hospital performance on the incentivised process and clinical measures is reported annually on the AQ website (http://www.advancingqualitynw.nhs.uk). We examine whether there is evidence that adoption of the scheme has translated into better health outcomes for patients. We evaluate the effect of AQ for only three of the five conditions. This means that our estimates of the effects of AQ are conservative, representing the lower bound of the actual effects as they do not take into account any benefits achieved in the remaining two clinical areas. Our cost estimates, however, do include the costs of the AQ programme as a whole, as it was not possible to separate out the costs applicable to each clinical area. The resulting estimates of cost-effectiveness are therefore also conservative, and at this stage assume that no health benefits were attained for hip and knee replacement and CABG patients.

3.3.6. Time horizon. We analyse the costs over the 18-month period from October 2008 to April 2010 and discount the effects over the expected patient life.

3.4. Costs of Advancing Quality

The total cost of the programme to commissioners was just over £13m, with only £5m of this consisting of the financial incentives (Table II). The ongoing running costs of £7m exceed the bonus payments and make up the majority of the costs. This result reinforces the importance of considering the costs of P4P beyond the incentive payments themselves. If, like five of the 14 previous studies identified in our earlier critique, we had failed to include any costs other than the bonuses paid out to the top performing hospitals, we would have underestimated the true cost of AQ by over 60%. Even if we exclude the set up costs, which it could be argued should be spread across a number of years, the incentive payments themselves still only represent 42% of the cost of the programme.

DOI: 10.1002/hec

Table II. Costs of the Advancing Quality programme

Activity	Costs
Set up costs	£990 000
Incentive payments	£5 054 489
Programme running costs	£7 015 531
One-off programme costs	£9844
Total costs	£13 069 864

Table III. Difference-in-differences estimates of Advancing Quality on percentage risk of mortality, readmission and days of hospital stay

	Mortality	Readmissions	LOS
Total incentivised	−0.9***	0.2	−0.3**
	[−1.4, −0.4]	[−0.3, 0.7]	[−0.5, −0.1]
AMI	−0.3	0.1	−0.3
	[−1.0, 0.4]	[−0.7, 0.9]	[−0.6, 0.1]
Heart failure	−0.3	0.7	−0.2
	[−1.2, 0.6]	[−0.4, 1.9]	[−0.6, 0.3]
Pneumonia	−1.6***	−0.0	−0.5**
	[−2.4, −0.8]	[−0.8, 0.7]	[−0.8, −0.1]

LOS, length of stay.
Between-region difference-in-differences estimates. *95%* confidence intervals in brackets, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Table IV. Difference-in-differences estimates of Advancing Quality on QALY tariffed mortality and cost-tariffed readmissions and days of hospital stay.

	Discounted QALYs	Readmissions £	LOS £
Total incentivised	0.07***	9.0	−62.4**
	[0.04, 0.11]	[−5.2, 23.3]	[−102.4, −22.3]
AMI	0.04	11.2	−58.1
	[−0.01, 0.10]	[−9.0, 31.4]	[−118.1, 2.0]
Heart failure	0.00	21.8	−31.6
	[−0.06, 0.07]	[−14.8, 58.3]	[−111.7, 48.4]
Pneumonia	0.13***	0.7	−82.1*
	[0.06, 0.19]	[−20.2, 21.6]	[−146.0, −18.2]

QALY, quality-adjusted life years; LOS, length of stay.
$*p < 0.05$, $**p < 0.01$, $***p < 0.001$. LOS costed at per diem HRG tariff. Readmissions costed at the HRG tariff of the readmission. QALYs estimated on the basis of life age and gender based healthy life expectancy for age at admission.

3.5. Effects of Advancing Quality

We estimate a statistically significant reduction in mortality and LOS associated with the introduction of AQ (Table III), which is statistically significant for pneumonia only when the three conditions are analysed individually. Readmission rates are unchanged. There are also statistically significant reductions in DANQALE and cost-tariffed LOS (Table IV), again statistically significant for pneumonia only when the three conditions are analysed separately.

DOI: 10.1002/hec

Table V. Total effects on outcomes in raw and tariffed units

	Total outcome changes in natural units			Total benefits/costs		
Condition	Mortality (deaths)	Readmissions	LOS (days)	ΔQALY	Readmissions £m	LOS £m
Total incentivised	−649	996	−22 802	5227	0.6	−4.4
AMI	−60	168	−4787	778	0.2	−1.1
Heart failure	−44	644	−2493	26	0.3	−0.5
Pneumonia	−580	−47	−16 540	4701	0.0	−3.0

LOS, length of stay costed at per diem HRG tariff. Readmissions costed at the HRG tariff of the readmission. QALYs estimated on the basis of life age and gender based healthy life expectancy for age at admission.

3.6. Cost-effectiveness of Advancing Quality

We estimate a reduction of 649 deaths[6] and a gain of 5227 QALYs as a result of the programme (Table V). At a QALY value of £20 000, this equals to an estimated health gain worth £105m.

Our estimates suggest that AQ resulted in 22 802 fewer days in hospital, saving £4.4m. Because of the structure of the payment system in operation in England, where payment for a hospital admission is largely independent of LOS, these cost savings would be claimed mostly by providers rather than commissioners. For readmissions, we estimate a statistically insignificant £0.6m increase in costs across all conditions.

4. DISCUSSION

P4P schemes are increasingly being used by purchasers as a means to encourage providers to improve their quality of care. Research to date has focused on whether such programmes induce changes on the targeted quality measures, commonly neglecting the more pertinent issue of their effect on health outcomes and costs. After critiquing the narrow range of costs and effects considered by previous evaluations, we developed an analytical framework to guide the future assessment of the cost-effectiveness of P4P programmes.

Our application of this framework to the AQ initiative reinforces the importance of considering costs beyond the incentive payments themselves, as failing to do so would have led us to include only 40% of the costs of the scheme from the commissioners' perspective. We have also estimated the incremental effects of AQ on mortality, readmissions and LOS directly, rather than relying on simulation modelling of the scheme's consequences. We observed statistically significant reductions in mortality and LOS attributable to the programme and converted the mortality reductions into expected QALY gains. Despite incorporating a wide range of programme costs into our evaluation, we still find it likely that AQ represented a cost-effective use of resources during the 18-month period under examination at standard UK threshold values. Crude estimates put the monetary value of the estimated QALYs gained at £105m, far exceeding the £13m spent by commissioners on the programme.

Some biases may be present in our analysis for a number of reasons, most of which lead us to underestimate rather than overestimate the benefits attributable to the AQ programme. Firstly, we are only able to estimate outcomes for three of the five incentivised conditions and therefore make the conservative assumption that no benefits are produced for hip and knee replacement or CABG patients. Secondly, we are unable to estimate any 'pure' quality of life effects not associated with mortality. Thirdly, we assume that any observed improvements in quality of care are transitory and will not affect future patients. However, the use of age-sex specific DANQALE estimates from the general population is likely to overestimate the health gains enjoyed by the additional survivors because the average life expectancy and health-related quality of life of individuals admitted to hospital for AMI, heart failure and pneumonia are likely to be lower than that of the general population.

[6]This figure equals that which would be produced by between-region DID estimation of Sutton *et al*. (2012) but is fewer than the 890 deaths arising from their triple-difference models.

DOI: 10.1002/hec

Nonetheless, just one QALY on average would need to be produced as a result of each death averted for AQ to be deemed cost-effective at the standard threshold of £20 000 for the value of a QALY.

We also estimated cost savings of £4.4m as a result of reduced LOS. Because of the structure of the payment system in operation, these cost savings would have been accrued to providers rather than payers. It is therefore rather puzzling that providers required financial incentives from purchasers to encourage such quality improving behaviour, when this behaviour is likely to have reduced their own costs. One possible explanation is that providers required the additional technological information on what represents best practice to realise such savings. Alternatively, the cost of providing the improved care may outweigh the reduced LOS cost savings, and so in the absence of the financial incentives it may not be efficient for providers to engage in quality improving behaviour.

Although it appears that AQ is likely to have represented a cost-effective use of resources during the 18-month period we evaluated, an important consideration for policy makers is its ability to continue generating improvements in the long run. This concern applies to all P4P schemes. It may be that P4P should be seen as a vehicle to kick start quality improving behaviours in the short term, which will then become engrained into routine. Alternatively, the observed improvements may simply represent transitory effort increases, which will fall away once the financial incentives are removed.

This is one of several aspects of P4P schemes about which there is little good quality evidence. These include: whether the incentives should be bonuses or fines; what size of incentive is required; whether payments should be made for outcomes or activities likely to lead to better outcomes; whether schemes should be tournaments or potentially reward all providers; and whether payment schedules should be linear or 'stepped'.

The intended and unintended behavioural responses of providers have formed the main focus of most research on P4P, not whether it is cost-effective. Yet resources spent on P4P also have opportunity costs. There are several P4P schemes in the health sector that would be worthy of cost-effectiveness analysis. We hope that the framework we have proposed will be developed further and applied to these schemes in the future.

APPENDIX A

TABLE A.1. PREVIOUS LITERATURE

Study no.	1	2	3	4	5	6	7
First author, year	Norton (1992)	Kouides *et al.* (1998)	Kahn *et al.* (2006)	Curtin *et al.* (2006)	Nahra *et al.* (2006)	Brown *et al.* (2007)	Parke (2007)
Country	USA	USA	USA	USA	USA	USA	USA
Setting	Nursing homes	Primary care clinics	Hospitals	Primary care clinics	Hospitals	Primary care clinics	Primary care clinics
Perspective	Health plan & providers	Health plan	Health plan	Health plan	Health plan	Providers	Health plan (& its enrolees)
Costs included:							
Development/set up costs	×	×	×	✓	×	?	×
Running costs	?	×	×	✓	✓	✓	✓
Treatment costs	✓	×	×	✓	×	✓	?
Incentive payments	✓	✓	✓	✓	✓	✓	✓
Outcomes:							
Incentivised measures	✓	✓	✓	×	✓	×	×
Intermediate	✓	×	×	×	×	×	×
Mortality	✓	×	×	×	✓	×	×
QALYs	×	×	×	×	✓	×	×
Non-incentivised care	×	×	×	×	×	×	×

DOI: 10.1002/hec

TABLE A.1. (Continued)

Study no.	8	9	10	11	12	13	14
First author, year	An *et al.* (2008)	Salize *et al.* (2009)	Rosenthal *et al.* (2009)	Ryan (2009)	Lee *et al.* (2010)	Sutton *et al.* (2010)	Walker *et al.* (2010)
Country	USA	Germany	USA	USA	China	UK	UK
Setting	Primary care clinics	Primary care clinics	Prenatal care	Hospitals	Primary care clinics	Primary care clinics	Primary care clinics
Perspective	Health plan	Health plan	Health plan	Health plan	National health insurance system	National health insurance system	National health insurance system
Costs included:							
Development/set up costs	✓	×	×	×	×	×	×
Running costs	✓	✓	×	×	×	×	×
Treatment costs	✓	✓	×	×	✓	×	✓
Incentive payments	✓	✓	✓	✓	✓	✓	✓
Outcomes:							
Incentivised measures	✓	✓	✓	×	✓	✓	✓
Intermediate	×	×	✓	×	✓	×	×
Mortality	×	×	×	✓	×	×	×
QALYs	×	×	×	×	×	×	✓
Non-incentivised care	×	×	×	×	×	✓	×

[1]QALYs, quality-adjusted life years.× = not reported, ? = unclear/lack of detail, ✓ = reported

ACKNOWLEDGEMENTS

This project was funded by the National Institute for Health Research Health Services and Delivery Research (NIHR HS&DR) programme (project number 08/1809/250). Visit the HS&DR website for more information. The views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR programme, NIHR, NHS or the Department of Health. We are grateful to Lesley Kitchen and Jane Harper of the Advancing Quality Alliance (http://www.advancingqualityalliance.nhs.uk/) for providing the data on the costs of the Advancing Quality programme and to participants in the UK Health Economists' Study Group in Oxford in June 2012 and a seminar at the University of York for their comments.

REFERENCES

An LC, Bluhm JH, Foldes SS, Alesci NL, Klatt CM, Center BA, Nersesian WS, Larson ME, Ahluwalia JS, Manley MW. 2008. A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. *Archives of Internal Medicine* **168**: 1993–1999.

Brown SES, Chin MH, Huang ES. 2007. Estimating costs of quality improvement for outpatient healthcare organisations: a practical methodology. *Quality & Safety in Health Care* **16**: 248–251.

Curtin K, Beckman H, Pankow G, Milillo Y, Greene RA. 2006. Return on investment in pay for performance: a diabetes case study. *Journal of Healthcare Management* **51**: 365–376.

Department of Health. 2008. Using the Commissioning for Quality and Innovation (CQUIN) payment framework.

Drummond MF, Jefferson TO. 1996. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* **313**: 275–283.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. 2005. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press: Oxford.
Emmert M, Eijkenaar F, Kemter H, Esslinger AS, Schoffski O. 2012. Economic evaluation of pay-for-performance in health care: a systematic review. *The European Journal of Health Economics* **13**: 755–767.
Greene SE, Nash DB. 2009. Pay for performance: an overview of the literature. *American Journal of Medical Quality* **24**: 140–163.
Holmstrom B, Milgrom P. 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* **7**: 24–52.
Kaarboe O, Siciliani L. 2011. Multi-tasking, quality and pay for performance. *Health Economics* **20**: 225–238.
Kahn CN, Ault T, Isenstein H, Potetz L, Van Gelder S. 2006. Snapshot of hospital quality reporting and pay-for-performance under medicare. *Health Affairs* **25**: 148–162.
Kelman S, Friedman JN. 2009. Performance improvement and performance dysfunction: an empirical examination of impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration, Research and Theory* **19**: 917–946.
Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH, LaForce FM. 1998. Performance-based physician reimbursement and influenza immunization rates in the elderly. The Primary-Care Physicians of Monroe County. *American Journal of Preventive Medicine* **14**: 89–95.
Lee TT, Cheng SH, Chen CC, Lai MS. 2010. A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. *The American Journal of Managed Care* **16**: 65–9.
Lester H, Schmittdiel 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.
Mason A, Walker S, Claxton K, Cookson R, Fenwick E, Sculpher M. 2008. The GMS Quality and Outcomes Framework: Are the Quality and Outcomes Framework (QOF) indicators a cost-effective use of NHS resources? *Joint Executive Summary*: *Reports to the Department of Health from the University of East Anglia & the University of York.*
Maynard A. 2012. The powers and pitfalls of payment for performance. *Health Economics* **21**: 3–12.
Mehrotra A, Damberg CL, Sorbero MES, Teleki SS. 2009. Pay for performance in the hospital setting: What is the state of the evidence? *American Journal of Medical Quality* **24**: 19–28.
Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JRC. 2006. Cost-effectiveness of hospital pay-for-performance incentives. *Medical Care Research and Review* **63**: 49S–72S.
National Institute for Health and Care Excellence (NICE). 2013. Guide to the methods of technology appraisal [Online]. Available: http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9/the-reference-case [Accessed 17.07.2013].
Norton EC. 1992. Incentive regulation of nursing homes. *Journal of Health Economics* **11**: 105–128.
Office for National Statistics. 2011. *England*, *Interim Life Tables*, *1980-82 to 2008-10* [Online]. Available: http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2008-2010/rft-ilt-eng-2008-10.xls [Accessed on 17.07.2013].
Parke DW, 2nd. 2007. Impact of a pay-for-performance intervention: financial analysis of a pilot program implementation and implications for ophthalmology (an American Ophthalmological Society thesis). *Transactions of the American Ophthalmological Society* **105**: 448–60.
Rosenthal MB, Li Z, Robertson AD, Milstein A. 2009. Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Services Research* **44**: 1465–1479.
Ryan AM. 2009. Effects of the premier hospital quality incentive demonstration on medicare patient mortality and cost. *Health Services Research* **44**: 821–842.
Salize HJ, Merkel S, Reinhard I, Twardella D, Mann K, Brenner H. 2009. Cost-effective primary care-based strategies to improve smoking cessation: more value for money. *Archives of Internal Medicine* **169**: 230–235.
Scott A, Sivey P, Ait Ouakrim D, Willenburg L, Naccarella L, Furler J, Young D. 2011. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrance Database of Systematic Reviews* **9**: CD008451.
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 Economics* **19**: 1–13.
Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. 2012. Reduced mortality with hospital pay for performance in England. *The New England Journal of Medicine* **367**: 1821–1828.
Walker S, Mason AR, 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. *British Journal of General Practice* **60**: e213–20.

DOI: 10.1002/hec