WORK-IN-PROGRESS: DO NOT CITE WITHOUT AUTHORS' PERMISSION

The cost-effectiveness of using financial incentives to improve provider quality: a framework and application

Rachel Meacock, Søren Rud Kristensen, Matt Sutton

Health Economics, Health Sciences Research Group, University of Manchester

Correspondence to: rachel.meacock@manchester.ac.uk

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 the UK. 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 programme costs. Nonetheless, our estimates of the value of the health gains produced in the first 18 months suggest that the scheme was a cost-effective use of resources. We conclude by noting the key areas of uncertainty and priorities for future data collection and research.

Paper presented to the Health Economists' Study Group, Oxford June 2012

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.

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 growing widespread adoption of P4P programmes, and much research by economists on this topic, there remains remarkably little evidence on the cost-effectiveness of such schemes.

A recent commentary on the topic by Maynard [1] highlights the 'curious' focus of research to date on the effectiveness of P4P schemes, with a neglect to their costs, and therefore cost-effectiveness. This gap in the evidence base is also noted by a number of reviews. Greene and Nash provide an overview of the literature on P4P published between 2004 and 2008 [2]. Of the 100 articles included in their annotated bibliography, only 3 are grouped under the heading of 'cost analysis' [3-5]. Mehrotra and colleagues systematically review the evidence on hospital-based P4P programmes, stating there to be approximately 40 such schemes targeted directly at inpatient care [6]. Despite this prevalence in the use of P4P, only 8 formal evaluations were found, covering just 3 different programmes. Of these 8 published studies, just 1 [4] attempted to estimate the cost-effectiveness of the programme in question.

Most recently, Emmert et al. present a systematic review of the literature concerning the economic evaluation of P4P [7], critically assessing the identified studies on their methodological quality according to the widely used Drummond and Jefferson checklist [8]. They identify 9 studies, of which only 3 are categorized as full economic evaluations [4, 9-10], and 6 as partial evaluations [3, 5, 11-14]. Of these 6, 4 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 [11-14]. The remaining 2 partial evaluations were simple cost comparisons, examining only the financial implications of the schemes in question [3, 5]. This comprehensive review concludes 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 offered by the Quality and Outcomes Framework (QOF) in the UK lead the Department of Health to commission a report in which a conceptual framework was developed to assess the cost-effectiveness of QOF indicators [15-16]. 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 substantial 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 the cost and benefits of incremental changes. Whilst 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 feel that it is necessary to take this a step further and consider whether P4P policies as a whole represent a cost-effective use of resources.

We therefore 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 Advancing Quality (AQ) scheme introduced in the North West of England in 2008 in an attempt to assess whether this P4P programme represents a cost-effective use of resources.

Methods

Critiquing previous evaluations

In light of the recently published Emmert et al. review [7], we thought it would be superfluous to appraise systematically the quality of the economic evaluations published to date. We instead 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 to ensure that no relevant articles published since their summary were missed. Studies known to the authors but not included in the Emmert et al. 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.

Developing the analytical framework

This review 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, as outlined in already established frameworks e.g. [17-18] in order to provide a more specific framework to guide the evaluation of P4P programmes.

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 CQUIN scheme. The programme aimed at improving quality in participating hospitals by paying for performance on 28 indicators across 5 conditions (see below). 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.

In assessing the effects of AQ we analyse mortality within 30 days of admission, length of stay (LOS) and readmissions for three of the five incentivised conditions (acute myocardial infarction (AMI), heart failure and pneumonia). We use patient level Hospital Episode Statistics (HES) data for patients admitted for one of the three AQ conditions under examination in the period 1st 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. In total our sample consists of 856,574 patients 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 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 OLS on

quarterly observations of risk-adjusted in-hospital mortality, LOS and readmission rates, 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

$$y_{jt} = \alpha_1 + u_j + v_t + \delta_{jt} D_j^1 \times D_t^2 + \varepsilon_{jt}$$

with y_{jt} being the outcome of interest (mortality rate, readmission rate or mean LOS) 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 takes take the value 1 if the hospital is located in the North West, and 0 otherwise, while the variable D_t^2 is 1 for all observations from the time period after the introduction of AQ and 0 before. Our main interest is in the coefficient on the interaction of these two variables δ_{jt}, which measures the mean difference in outcome at AQ hospitals compared to hospitals in the rest of England in the 18 months we observe AQ in effect, controlling for fixed hospital and time effects. The main effects of D_j^1 and D_t^2 are not included in the regression, as they are perfectly collinear with the time and hospital fixed effects.

To evaluate the cost-effectiveness of AQ, we next assign monetary values to the estimated outcomes. For mortality, we consider potential gains in terms of quality adjusted life years (QALYs), taking into account the average age and gender of the patients treated for each condition. This is similar to the approach suggested by [19], although currently we do not discount the QALY gains as otherwise suggested.

We thus base our estimates of gained life years on the life expectancy at the average age and gender for patients subject to AQ. We use the 2008-10 Interim Life Tables (ONS) as a source for life expectancy [20]. To adjust for quality of life we use the proportion of life expectancy at age 65 expected to be spent in good health for the English population from [21] and assume quality adjustments of 1 for years in full health and 0.5 for years not at full health.

With respect to valuing the changes in LOS, ideally, we would value these according to the appropriate HRGs in order to estimate the resulting cost savings, but at this time we are only able to produce a crude estimate using the mean per diem tariff for days above the trim point across all HRGs in 2009/10 tariffs of £227 per day. Likewise, the change in readmissions would ideally be valued at the mean value of the HRGs in which readmissions after discharge

from the three conditions fall, but we use as an equally crude estimate of the mean value of all HRGs in tariff system 2009/10 of £3,036.

Results

Previous evaluations

We identified 14 studies examining the cost of P4P schemes, the details of which are presented in table A1 in the appendix. The majority of these schemes were operated in the US [3-5, 9-10, 12-14, 22-23], with 2 in the UK [15, 24], and 1 in each of Germany [25] and China [11]. The most common setting for the programmes under evaluation were primary care clinics [3, 5, 9-11, 15, 23-25], followed by hospitals [4, 14, 22]. 9 of the 10 US evaluations were undertaken from the perspective of the health plan [3-5, 9-10, 12-14, 22], with 1 extending this to include the plan's enrolees [5] and another also considering the providers of the incentivised treatment [12]. Just 1 evaluation was performed purely from the providers' perspective [25], and the remaining 3 from that of government-run health systems [11, 15, 24]. The range of costs included by many of the studies were however inconsistent with their stated perspectives, often failing to encompass all of the relevant cost categories. Just 2 evaluations clearly incorporated the costs associated with the development and set-up of the P4P schemes in question [3, 9], and only 6 included the ongoing running costs [3-5, 9, 23, 25]. 7 studies made some attempt to measure the increased costs associated with providing the incentivised treatments [3, 9, 11-12, 15, 23, 25], whilst 5 failed to even consider any costs beyond the incentive payments themselves [10, 13-14, 22, 24].

Of the 14 studies examining the cost of P4P programmes, 11 also made some attempt to estimate the effects of the scheme in question [4, 9-15, 22, 24-25]. The range of effects considered by these 11 studies were however narrow. The incentivised performance measures were by far the most commonly used metric of effect, with all but 1 evaluation reporting results on these process or clinical measures [14]. 4 evaluations failed to consider anything beyond these incentivised measures [9-10, 22, 25], making no attempt to link any quality improvements achieved to health outcomes. 3 studies examined intermediate outcomes such as hospitalisations and lengths of stay [11-13], and 3 report reductions in mortality [4, 12, 14]. Just 2 of the evaluations identified attempted to express the effects of P4P schemes in terms of QALYs [4, 15], and only 1 looked at the potential effects on non-incentivised areas of care [24].

The omission of relevant cost categories by many previously conducted evaluations, along with the lack of evidence regarding the effects of P4P in terms of health outcomes means that conclusions regarding the value for money of the programmes in question cannot be made. We now set out our analytical framework, which aims to guide the future assessment of the cost-effectiveness of P4P programmes in order to enable such conclusions to be drawn.

Analytical framework

1 Perspective

The relevant perspective will vary by evaluation, and depend upon the particulars of the P4P programme under examination. In the UK this is often likely to, although will not always be, that of the National Health Service (NHS) and personal social services (PSS), consistent with the perspective specified by the National Institute for Health and Clinical Excellence (NICE) in their reference case [18]. Whatever the chosen viewpoint for the analysis, this 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/tax payers, 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 the P4P scheme and perform the tasks necessary to improve performance on the stipulated quality measures. 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. Whilst some/all of these costs may be offset by the incentive payments, there is no guarantee that providers will actually receive the 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 (called withholds in the US literature) if performance benchmarks are not met. It is therefore important to consider the costs and effects of P4P schemes for those providing the incentivised care as well as those paying to incentivise it.

2 Comparator

A clear comparator is essential for any economic evaluation [17], representing what would have happened in the absence of the programme under examination. The relevant counterfactual will again depend upon the P4P scheme under evaluation, as well as the perspective taken. An important consideration when evaluating P4P schemes is whether to compare to same additional resources but paid in a different manner or whether to compare to no bonus payments at all. 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. 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 however vital 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 confounding factors such as general time trends which may have also affected the outcomes under examination.

3 Cost categories

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

- Set up/development costs – e.g. staff time, investment in infrastructure. These costs can be spread across the expected life time 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, pharmacy. The perspective of the evaluation will dictate whether these costs are relevant (see 1), but it is always worth noting that providing the incentivised treatments will have cost consequences for health care providers.
- Cost savings – e.g. reduced complications, lengths of stay, readmissions and lifetime costs. It is assumed that improving the quality of care will produce superior health outcomes, which in turn has consequences for future health care 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 above cost categories and examples given for each are not intended to be exhaustive, but instead act as a guide to 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 against the resources involved in accurately estimating it when considering the precision with which we must estimate. 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.

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 is characterised by the possible health gains forsaken 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 off and retained to fund the incentive scheme, or the duties of existing members of staff may be changed to focus on the areas of care incentivised. Whilst this does not involve any additional spending, these resources still have an opportunity cost in terms of care displaced, and attempts should be made to account for this.

5 Outcomes

The main outcomes recorded for P4P programmes are the targeted quality measures upon which performance is judged, which vary substantially depending on the scheme. If these are process rather than clinical outcome measures, then evidence on their link with health outcomes (i.e. length and quality of life) should be presented if available. These measures of effects can also be supplemented with data covering outcomes such as reduced readmissions and mortality, depending on what is relevant to the scheme and interventions being targeted. It can be difficult to attach value to the measures collected in order to judge whether the resulting quality improvements were in fact worth the money spent on P4P. Ideally, benefits would be expressed in terms of quality-adjusted life years (QALYs) in order to permit comparison with standard cost-effectiveness thresholds in the UK [15, 18] .

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 when quality is multidimensional. If incentives divert the existing efforts of providers away from un-incentivised areas of care rather than simply promoting additional effort in the targeted areas, this could result in unintended consequences for patients [26]. 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 un-incentivised dimensions are substitutes or complements to the agent, it may even be undesirable to pay for performance [27-28]. 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 full capture their effects, both intended and unintended.

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 upon year. Whilst 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 [29]. Alternatively, incentivised behaviours may have become routine and therefore continue even after payments are withdrawn.

Advancing Quality (AQ)

We first work through the discussion points raised by the framework in relation to the evaluation of AQ, before presenting estimates of the scheme's costs and effects.

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. In future work we are also planning to survey a sample of the trusts involved in the scheme in order to estimate any costs which they may have incurred as a result of participating in AQ. As the programme ran under a tournament system only half of the hospitals involved actually received bonus payments at each payout, meaning that providers may have incurred substantial participation costs yet received no financial rewards for their efforts.

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 Cost categories

We go beyond the incentive payments themselves and attempt to include all of the relevant costs incurred by commissioners as a result of AQ. The set-up costs represent one-off lump-sum grants given to providers, designed to cover things such as investments in infrastructure necessary to enable the required data collection. As AQ was merged with

another national P4P policy 18months after its introduction, and the expected life time of this new policy is unknown, the set up costs are not apportioned but instead assessed at the time at which they were incurred. This assumption therefore means that our estimate of the cost (and so cost-effectiveness) of AQ over the first 18months for which it ran represent the upper bound of the actual costs applicable to this period.

The financial incentives paid out to providers are then presented, along with the ongoing running costs of the programme and any other one-off costs incurred within the period of assessment. In the case of AQ the general running costs of the programme encompass things such as 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 various other procurements. Finally we examine the potential cost savings resulting from reduced length of stay and readmissions, and discuss who the residual claimant on any such savings may be.

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 are unable to determine exactly what this money would have been spent on in the absence of the policy, but acknowledge that there are likely to have been possible health gains forsaken from alternative uses of those resources. Whilst we cannot capture the true opportunity cost of the care displaced, the use of the standard UK cost-effectiveness thresholds can be indicative of this.

5 Outcomes

Hospital performance on the incentivised process and clinical measures is reported annually on the AQ website (http://www.advancingqualitynw.nhs.uk) and so not examined here. We instead examine whether there is evidence that adoption of the scheme has translated into superior health outcomes for patients. We evaluate the effect of AQ on mortality, LOS and readmissions for 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 on mortality, LOS and readmissions as they do not take into account any benefits achieved in the remaining 2 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 and coronary artery grafting surgery patients.

6 Time horizon

We analyse the costs and effects of AQ over the 18month period from October 2008 to April 2010 inclusive, and do not discount over this short period. We do however intend to incorporate discounting into our assessment of the possible future QALY gains resulting from the observed reductions in mortality in future work, but are unable to do so at this time.

Costs of AQ

The costs of AQ to commissioners over the 18month period under examination are shown in Table 1 broken down into the various cost categories. The total cost of the programme was just over £13million, with only £5million of this consisting of the financial incentives. The ongoing running costs of the scheme actually exceed the bonus payments, making up the majority of the costs at just over £7million. This result reinforces the importance of considering the costs of P4P beyond the incentive payments themselves. If, like 5 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.

Table 1: Costs of AQ

Activity	Costs
Set up costs	£990,000
Incentive payments	£5,054,489
Programme running costs	£7,015,531
One-off programme costs	£9,844
Total costs	**£13,069,864**

Effects of AQ

The effects of AQ on mortality, LOS and readmissions are presented in Table 2. The results are presented for the three conditions combined and for each condition separately. We observe an overall statistically significant reduction in mortality and LOS associated with the introduction of AQ, which is statistically significant for pneumonia only when the three conditions are analysed individually. Readmission rates are unchanged overall and for pneumonia and heart failure, but increased for AMI patients.

Cost-effectiveness of AQ

To assess the cost-effectiveness of AQ, we now assign monetary values to the estimated changes in outcomes. The results are presented in Table 3. We estimate a gain of 6689 QALYs as a result of the reduction in mortality for the programme as a whole. At a QALY value of £20,000 this equals an estimated health gain worth £134 million (£191 million if a QALY is valued at £30,000).

Our estimates suggest that AQ saved 22684 in-hospital days in the period under study. This estimates a £5 million reduction for the period we observe. Due to the structure of the payment system in operation in the UK, these cost savings would have largely been claimed by providers rather than commissioners. For readmissions, we estimate a statistically insignificant £0.5 million increase in costs across all conditions.

Due to the uncertainty around the methods used to estimate the potential QALYs gained as a result of the estimated reductions in mortality, we calculated the number of additional QALYs that it would be necessary to gain as a result of AQ in order for the programme to be deemed cost-effective at the standard UK cost-effectiveness threshold. Table 4 shows the results of these calculations, and compares these to the estimated number of deaths averted in order to show how many QALYs would need to be gained for each averted death in order for AQ to represent value for money at the standard cost-effectiveness thresholds. We estimate that if just 1 QALY were gained for each death that was averted, AQ would be deemed cost-effective if a QALY is valued at £20,000. Using the upper threshold of £30,000 for the value of a QALY, it would be necessary to gain just 0.67 QALYs from each death averted.

Table 2: Estimated effects of AQ on mortality, LOS and readmissions

	Mortality	LOS	Readmissions
Total incentivized	-0.915***	-0.321**	0.313
	(0.000)	(0.003)	(0.142)
AMI[a]	-0.307	-0.254	1.248**
	(0.406)	(0.104)	(0.004)
Heart failure	-0.288	-0.157	-0.341
	(0.535)	(0.469)	(0.427)
Pneumonia	-1.591***	-0.454**	0.073
	(0.000)	(0.008)	(0.800)

Notes: Between region difference-in-differences estimates. N= 5443 for the pooled regression, N=1815 for AMI and N=1814 for heart failure and pneumonia.
p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
a) AMI is Acute Myocardial infarction

Table 3: Estimated value of outcomes on mortality, LOS and readmissions under AQ

	Patient characteristics			Mortality		LOS		Readmissions	
Condition	Patients	Avg. age	Prop. Males	ΔQUALY	£m	ΔDays	£m	ΔReadmissions	£m
Total incentivised	70666	73	0.55	6689	133.77	-22684	-£5.15	154	0.47
Acute Myocardial Infarction	18762	70	0.62	686	13.73	-4766	-£1.08	234	0.71
Heart Failure	15476	77	0.53	392	7.84	-2430	-£0.55	-53	-0.16
Pneumonia	36428	72	0.50	6369	£127.37	-16538	-£3.75	27	£0.08

Notes: Estimates assuming value per QUALY is £20,000, value per day saved is 227 (average per diem tariff for days above trim point across all HRGs in 2009/10, value per readmission is 3,036 (mean non-eletive spell tariff across all HRGs 2009/10). QUALYs estimated on the basis of life expectancy at average age and gender of patients and healthy years expectancy at 65 and assuming a value of unhealthy years of .05.

Table 4: Cost effectiveness of AQ

Value of a QALY	QALYs needed	QALYs needed per death averted
£20,000	653.49	1.01
£30,000	435.66	0.67

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 health care 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 caused us to only capture 40% of the true costs of the scheme from the commissioners' perspective. We were also able to directly estimate the incremental effects of AQ in terms of LOS, readmissions and reduced mortality, rather than relying on simulation modelling of the scheme's consequences. We observed statistically significant reductions in LOS and mortality attributable to the programme, and attempted to convert these mortality reductions into potential QALY gains taking into account the average age and gender of the patients treated for each condition. 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 at standard UK threshold values. Crude estimates put the monetary value of the estimated QALYS gained at £134million, which far exceeds the £13million spent by commissioners on the programme. Note that this estimated stream of QALY gains has not been discounted, but this is something we intend to investigate in future work. Approaching the problem from a different angle, just 1 QALY 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 examine the potential for cost savings as a result of reduced LOS, our crude estimates of which suggest that £5million in cost savings could have been made. Due to 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.

Whilst it appears that AQ is likely to have represented a cost-effective use of resources during the 18month period for which we evaluated, an important consideration for policy makers is its ability to continue generating improvements in quality and health outcomes 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.

The work presented here is work in progress and we acknowledge its various limitations. Reliance on administrative data means that were forced to use proxies for costs and benefits collected for all patients rather than the intensity of care received by individuals. We were also unable to present any effects for the other two clinical areas covered by AQ but hope to do so at a later date.

Despite these limitations, we feel that our evaluation represents real progress in terms of the range of costs and effects considered and the methodology employed. Our intention was not to dictate a rigid set of rules, but to instead highlight the various issues that we feel are often overlooked when evaluating P4P programmes. We hope that this framework will open a dialog and result in more thorough analysis of future schemes.

There are 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’ like target payments. The intended and unintended behavioural responses of providers have been the focus of most research on this topic and not whether P4P is cost-effective. There are several P4P schemes in the UK health sector that would be worthy of cost-effectiveness analysis. Many, such as CQUIN and

the non-payment re-admissions policy, operate using fines rather than bonuses. These schemes, which require providers to make quality improvements or lose revenue, would require us to develop cost-effectiveness frameworks that are more imaginative in the wider, unintended consequences that they consider.

References

1. **Maynard, A., *The powers and pitfalls of payment for performance.* Health Economics, 2012. 21(1): p. 3-12.**
2. **Greene, S.E. and D.B. Nash, *Pay for Performance: An Overview of the Literature.* American Journal of Medical Quality, 2008.**
3. **Curtin, K., et al., *Return on Investment in Pay for Performance: A Diabetes Case Study.* Journal of Healthcare Management, 2006. 51(6): p. 365-376.**
4. **Nahra, T.A., et al., *Cost-Effectiveness of Hospital Pay-for-Performance Incentives.* Medical Care Research and Review, 2006. 63(1 suppl): p. 49S-72S.**
5. **Parke, D.W., 2nd, *Impact of a pay-for-performance intervention: financial analysis of a pilot program implementation and implications for ophthalmology (an American Ophthalmological Society thesis).* Trans Am Ophthalmol Soc, 2007. 105: p. 448-60.**
6. **Mehrotra, A., et al., *Pay for Performance in the Hospital Setting: What Is the State of the Evidence?* American Journal of Medical Quality, 2009. 24(1): p. 19-28.**
7. **Emmert, M., et al., *Economic evaluation of pay-for-performance in health care: a systematic review.* Eur J Health Econ, 2011.**
8. **Drummond, M.F. and T.O. Jefferson, *Guidelines for authors and peer reviewers of economic submissions to the BMJ.* BMJ, 1996. 313(7052): p. 275-283.**
9. **An, L.C., et al., *A Randomized Trial of a Pay-for-Performance Program Targeting Clinician Referral to a State Tobacco Quitline.* Arch Intern Med, 2008. 168(18): p. 1993-1999.**
10. **Kouides, R.W., et al., *Performance-based physician reimbursement and influenza immunization rates in the elderly.* American Journal of Preventive Medicine, 1998. 14(2): p. 89-95.**
11. **Lee, T.T., et al., *A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment.* Am J Manag Care, 2010. 16(1): p. 65-9.**
12. **Norton, E.C., *Incentive regulation of nursing homes.* Journal of Health Economics, 1992. 11(2): p. 105-128.**
13. **Rosenthal, M.B., et al., *Impact of Financial Incentives for Prenatal Care on Birth Outcomes and Spending.* Health Services Research, 2009. 44(5p1): p. 1465-1479.**
14. **Ryan, A.M., *Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost.* Health Services Research, 2009. 44(3): p. 821-842.**
15. **Walker, S., et al., *Value for money and the Quality and Outcomes Framework in primary care in the UK NHS.* Br J Gen Pract, 2010. 60(574): p. e213-20.**
16. **Mason, A., et al., *The GMS Quality and Outcomes Framework: Are the Quality And Outcomes Framework (QOF) Indicators a Cost-Effective Use of NHS Resources?*, in *Joint Executive Summary: Reports to the Department of Health from the University of East Anglia & the University of York*. 2008.**
17. **Drummond, M.F., et al., *Methods for the Economic Evaluation of Health Care Programmes.* 3rd ed. 2005, Oxford: Oxford University Press.**

18. National Institute for Health and Clinical Excellence (NICE), *Guide to the methods of technology appraisal*. 2008: London.
19. Dawson, D., et al., *Developing new approaches to measuring NHS outputs and productivity*, in *CHE Research Paper 6*. 2005, Centre for Health Economics, University of York.
20. Office for National Statistics. *England, Interim Life Tables, 1980-82 to 2008-10*. 2011; Available from: http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2008-2010/rft-ilt-eng-2008-10.xls.
21. Office for National Statistics. *Health expectancies in the United Kingdom 2000-2002 to 2007-2009*. 2011; Available from: http://www.ons.gov.uk/ons/rel/disability-and-health-measurement/health-expectancies-at-birth-and-age-65-in-the-united-kingdom/2007-09/rt---he-2000-2009.xls.
22. Kahn, C.N., et al., *Snapshot Of Hospital Quality Reporting And Pay-For-Performance Under Medicare.* Health Affairs, 2006. 25(1): p. 148-162.
23. Brown, S.E.S., M.H. Chin, and E.S. Huang, *Estimating costs of quality improvement for outpatient healthcare organisations: a practical methodology.* Quality and Safety in Health Care, 2007. 16(4): p. 248-251.
24. Sutton, M., et al., *Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers.* Health Economics, 2010. 19(1): p. 1-13.
25. Salize, H.J., et al., *Cost-effective Primary Care-Based Strategies to Improve Smoking Cessation: More Value for Money.* Arch Intern Med, 2009. 169(3): p. 230-235.
26. Kelman, S. and J.N. Friedman, *Performance Improvement and Performance Dysfunction: An Empirical Examination of Impacts of the Emergency Room Wait-Time Target in the English National Health Service.* SSRN eLibrary, 2007.
27. Kaarboe, O. and L. Siciliani, *Multi-tasking, quality and pay for performance.* Health Economics, 2011. 20(2): p. 225-238.
28. Holmstrom, B. and P. Milgrom, *Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.* Journal of Law, Economics, & Organization, 1991. 7(ArticleType: research-article / Issue Title: Special Issue: [Papers from the Conference on the New Science of Organization, January 1991] / Full publication date: 1991 / Copyright © 1991 Oxford University Press): p. 24-52.
29. Lester, H., et al., *The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators.* BMJ, 2010. 340.

Table A1: Previous literature

First author, year	Norton (1992)	Kouides (1998)	Kahn (2006)	Curtin (2006)	Nahra (2006)	Brown (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 enrollees)
Costs included:							
Development/set up costs	X	X	X	✓	X	?	X
Running costs	?	X	X	✓	✓	✓	✓
Treatment costs	✓	X	X	✓	X	✓	?
Incentive payments	✓	✓	✓	✓	✓	✓	✓
Outcomes:							
Incentivised measures	✓	✓	✓	X	✓	X	X
Intermeidiate	✓	X	X	X	X	X	X
Mortality	✓	X	X	X	✓	X	X
QALYS	X	X	X	X	✓	X	X
Non-incentivised care	X	X	X	X	X	X	X

First author, year	An (2008)	Salize (2009)	Rosenthal (2009)	Ryan (2009)	Lee (2010)	Sutton (2010)	Walker (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	✓	X	X	X	X	X	X
Running costs	✓	✓	X	X	X	X	X
Treatment costs	✓	✓	X	X	✓	X	✓
Incentive payments	✓	✓	✓	✓	✓	✓	✓
Outcomes:							
Incentivised measures	✓	✓	✓	X	✓	✓	✓
Intermeidiate	X	X	✓	X	✓	X	X
Mortality	X	X	X	✓	X	X	X
QALYS	X	X	X	X	X	X	✓
Non-incentivised care	X	X	X	X	X	✓	X

Notes: X=not reported, ?=unclear/lack of detail, ✓=reported