

# Evaluation of the Commissioning for Quality and Innovation Framework

## Final Report

---

February 2013

### *Authors*

*University of Nottingham*

*University of Manchester*

*Ruth McDonald*

*Søren Rud Kristensen*

*Sabeeh Zaidi*

*Matthew Sutton*

*Sarah Todd*

*Frederick Konteh*

*Kasser Hussein*

*Sue Brown*

This project was funded by the Department of Health Policy Research Programme, with additional support from the NIHR Health Services and Delivery Research Programme. The views expressed in this report are those of the authors and do not necessarily reflect those of the DH or NIHR.

## Table of Contents

Introduction .....	2
Methods.....	5
Analysis of the structure and content of the 2010/11 schemes.....	9
Qualitative analysis of the CQUIN process .....	13
Quantitative analysis of the impact of CQUIN goals.....	20
Conclusion from CQUIN evaluation .....	21
Findings from our evaluation of other relevant financial incentive initiatives.....	24
Lessons from CQUIN, AQ and BPT for the design of financial incentive initiatives .....	28
Tables .....	32
References .....	40

## **Introduction**

Aligning financial incentives with health system goals is a key aim of policy makers in the UK and beyond. In the NHS, recent changes have been made to the way that GPs, pharmacists and dentists are remunerated in an attempt to achieve such alignment. In 2009 the Commissioning for Quality and Innovation (CQUIN) payment framework was introduced to cover other aspects of care for NHS patients. CQUIN makes a proportion of income conditional on the achievement of locally agreed quality improvement and innovation goals for providers of acute, ambulance, community, mental health and learning disability services. The CQUIN framework has undergone various changes since April 2009 and it has been implemented against a background of considerable change in the NHS more generally.

This report presents findings from the evaluation of CQUIN commissioned by the Department of Health. The principal aim is to provide lessons to inform future policy-making in this area.

### **The research context**

The increased adoption of financial incentives to improve quality is occurring despite a somewhat mixed evidence base. Although the number of published evaluations of incentive schemes is growing relatively rapidly, the scientific quality of these is limited, due to the non-randomised nature of the studies<sup>1</sup>. A recent Cochrane review found no evidence that financial incentives aimed at changing practice in healthcare settings improve patient outcomes<sup>2</sup>. However, our evaluation of Advancing Quality in the NHS North West<sup>3</sup>, found that the scheme was associated with reduced mortality, although incentives were combined with other mechanisms including public reporting. Participation in the initiative was voluntary, but all eligible organisations participated. A recent review of published randomised controlled trials reported that the effectiveness of incentive programmes is 'highly variable' but that such programmes can improve care<sup>1</sup>. This found no clear relationship between incentive size and impact. It also suggested that incentivised targets should be based on baseline 'room for improvement' to avoid self-selection of high performers and 'ceiling effects'. The CQUIN Impact Assessment prepared as part of the process of developing and implement the CQUIN policy drew on available research evidence to inform the design of the CQUIN framework. This is discussed in more detail in the methods section of this report.

### **The changing policy context**

The CQUIN policy underwent a number of changes during the period covered by the evaluation. The size of the incentive in the CQUIN scheme was increased from 0.5% to 1.5% after the first year of introduction. (It was subsequently increased to 2.5% from April 2012). Additionally, despite the emphasis on local negotiations and ownership, two national goals

were introduced for acute providers in 2010/11 (and other national goals have since been introduced). For non-acute providers, participation in CQUIN was voluntary in Year 1, but this became mandatory from April 2010. In addition, guidance for 2010/11 emphasised stretch targets and outlawed payment for data collection (as opposed to improved performance) only. However, this was subsequently relaxed in recognition of the need to, at times, reward data collection effort as part of the first phase of quality improvement in some areas.

In addition, the Transforming Community Services (TCS) agenda which required PCTs to divest themselves of their community provider functions resulted in changed relationships between commissioners and their newly independent community providers. Furthermore, the development of clinical commissioning groups and related PCT restructuring resulted in high staff turnover, which impacted on the stability of working relationships between commissioners and providers.

## **Research Objectives**

The Department of Health call for research was aimed at providing:

- A national CQUIN picture containing structured qualitative and quantitative information regarding the content, goals, processes of agreement and changes in these over time to permit systematic long run assessment and monitoring of CQUIN schemes in England.
- Recommendations on further data to collect to maintain a national dataset and enable evaluation of the full impact of the CQUIN payment framework.
- A multi-method evaluation of the impact of CQUIN on behavioural change, processes and relationships and observed and hypothesised impacts (desired and otherwise) on quality and the wider NHS.

## Methods

The overarching philosophy underpinning the approach was one of Realistic Evaluation<sup>4</sup>. This views policies as 'theories incarnate'. In other words, whenever a policy is implemented, it is testing a theory about change. Realistic evaluation starts by making explicit the theories about how the policy might work, identifying the mechanisms through which specific outcomes are hypothesised to occur. In addition realistic evaluation recognises that the same policy can result in different outcomes in different places. It is important, therefore to collect and analyse data on implementation in a number of contexts, in order to identify the contextual factors which have an impact on policy outcomes.

### Contexts, mechanisms and outcomes

A number of theories of cause and effect were contained within the CQUIN Impact Assessment prepared as part of the process of developing and implement the CQUIN policy. CQUIN is intended to support a 'cultural shift by embedding quality improvement and innovation as part of the commissioner-provider discussion everywhere through making the payment system reflect quality' according to the Department of Health's Impact Assessment<sup>5</sup>. Although a nationally mandated scheme was considered within the DH Impact Assessment (IA), a scheme which encompassed mainly local goals was seen as preferable since this would 'ensure provider involvement in developing schemes and ....build on local initiatives and enthusiasm using standard metrics'. This emphasis on local goals and provider involvement chimes with the sentiments contained in *High Quality Care for All*, which proposed that clinical teams should use clinical indicators to measure the quality of care they provide, highlight areas for improvement and track the changes they deliver.

CQUIN is also aimed at increasing the focus of Boards on quality and innovation. Whilst the IA recognizes the importance of front line clinicians and clinical teams, it suggests that institutionally, 'there is little to systematically reinforce a focus on quality, and as a result, the extent of the emphasis on quality improvement and innovation varies across teams and health economies. In particular, Boards are sometimes not fully engaged'. The IA contrasts the Board's focus on activity and finance with matters of quality, suggesting that making 'quality improvement goals a mandatory part of contracts will avoid quality being crowded out of discussions. Attaching a small amount of money will provide an explicit signal and incentive to ensure that quality is an essential part of conversations and negotiations at Board level, which is likely to increase the institutional focus upon quality and hence improve outcomes for patients in the longer term'. Attaching money to quality is seen as 'likely to have a significant impact on increasing the focus on quality improvement and innovation in contract discussions if there is local agreement and enthusiasm around the priority areas of quality improvement'.

Evidence suggests that financial incentives need to be sufficiently large to engage providers, however, the percentage of total remuneration required to change behaviour at the organisation (as opposed to individual) level is relatively low, with estimates suggesting

between 1 and 4% is sufficient. The increase in the size of the CQUIN scheme from 0.5% to 1.5% after the first year of introduction can be seen as in line with this evidence.

In addition to increasing the focus of Boards on quality as part of a process of improving care within institutions, CQUIN is also intended to encourage a focus on quality and innovation 'across pathways'. Whereas the literature on performance measurement highlights the importance of choosing measures which are within the control of those whose performance is being measured, the phrase 'across pathways' might be interpreted as intending commissioners to work with a range of providers in order to encourage joint working (as opposed to for example, cost shifting and silo thinking) across these provider interfaces.

In summary then, the policy intent is that attaching money to quality and innovation, will focus Board members' attention, bring commissioners and providers together to increase dialogue and allow providers to contribute to the content of CQUIN goals. In addition, local flexibility for agreeing CQUIN goals is intended to capitalise on agreement and enthusiasm about local priority areas (in contrast to having goals imposed in a 'top down' fashion). The emphasis on dialogue between commissioners and providers across pathways, suggests that CQUIN should encourage joint working across providers. Furthermore, CQUIN is intended to achieve a 'cultural shift'. In some senses this is a policy outcome, but, it is also a mechanism, since a shift in culture is intended to lead to improved quality of care. This shift in culture might be hypothesised to occur as a result of this increased focus on quality, but also as a consequence of closer joint working, leading to shared objectives and improved relationships between commissioners and providers.

In order to evaluate CQUIN we undertook a mixed-methods study. We aimed to assess the impact of CQUIN on quality nationally using quality and quantitative methods, as well as categorising schemes as part of the national picture analysis.

### **National picture analysis**

To describe the national picture, we obtained copies of every CQUIN scheme from all but a handful of providers and classified these according to headings agreed with the Department of Health CQUIN team in order to provide a national, structured picture of CQUIN schemes.

Before undertaking the quantitative analysis of the structure and content of the 2010/11 CQUIN schemes we cleaned the data for differences in spelling and obvious misclassifications, and aggregated goals that had very similar content. For example, we combined the goals relating to 'patient satisfaction and 'patient involvement' because, although these goals are distinct in principle, commissioners had used the same indicators to operationalise these goals. To assess the size and complexity of the CQUIN schemes we counted the number of goals and indicators in each of the four sectors. For each sector, we then identified the 10 goals that occurred most frequently in local CQUIN schemes to identify which goals were perceived by commissioners and providers as most important. As another indicator of importance, for each sector we calculated the total weight attached to the indicators making up the goals. We then identified all of the indicators used for one

commonly-included goal (discharge planning) to examine the amount of variation in local indicator development.

Using the 2010/11 national picture, we also analysed the proportion of local indicators classified by scheme developers as incentivising the quality domains of structure/process, outcome, action plan or data collection. We also assess the proportion of indicators classified by developers as covering each of the quality dimensions: safety, effectiveness, patient experience and innovation. According to the CQUIN guidance, each CQUIN scheme should have contained at least one indicator in each domain. We investigated for what proportion of schemes this requirement was fulfilled and what proportions of schemes lacked an indicator in each of the four domains.

We also examine whether baseline values were specified in local schemes and, whether scheme developers used the option of paying for data collection where the baseline value had yet to be collected.

### **Qualitative analysis of process**

For our qualitative analysis, we used the theories of change underpinning CQUIN to develop sensitising concepts<sup>6</sup> to provide a general sense of reference and guidance in approaching the empirical data.

We explored the processes surrounding CQUIN using observation of meetings (373 comprising 800 hours of observation) in 12 case study sites, as well as 230 formal interviews with NHS staff from those sites. In addition, less formal conversations were conducted with NHS staff immediately prior to or just after meetings.

Of the formal interviews, 69 were from acute providers, 43 from community providers, 65 from mental health providers and 53 from commissioner organisations. 55 staff at director or deputy director level were interviewed and 128 managers, although the latter covered a wide range of roles from finance and administrative staff to clinical managers (mainly nurses). Our approach was to observe meetings between commissioners and providers at which CQUIN was discussed. These discussions were usually part of a wider meeting to discuss performance against contract more generally. Attending these meetings on a regular basis enabled us to track the process of CQUIN scheme development and negotiation, implementation, monitoring and final payment. Combined with interviews, this approach also enabled us to compare the stated intentions of interviewees with what happened in practice.

One of the drawbacks with this approach was that front line staff were usually not present at these meetings. We used a combination of asking meeting attendees for contacts and support from NIHR clinical research network staff, to identify and recruit front line clinicians.

Furthermore, since we were undertaking a parallel study looking at Best Practice Tariffs, we also drew on findings from interviews conducted for that study (93 of which were outside of case study sites) to enable us to examine CQUIN in the context of other levers for change available to commissioners and to draw on experiences from beyond our 12 case study sites.

A case study was defined as a health economy comprising a commissioner (including the related specialised commissioning input) and the providers from whom it commissioned care.

All interviews were digitally recorded and transcribed verbatim. Analysis initially involved coding transcripts using NVivo software and identifying themes. A constant comparative method was used to interpret the data<sup>7</sup>. Key concepts were identified using an open coding method. Once coding was complete, the codes that had common elements were merged to form categories. Analysis moved from within-case (i.e. a single health economy) to cross-case analysis to identify both site specific and more general issues.

### **Quantitative analysis of impact**

We focus our assessment of impact on the defining feature of CQUIN, the local content of schemes rather than the national or regional indicators. To assess whether the locally developed schemes were successful in achieving the desired outcome, we focus on whether Trusts that had specific topics as a goal in their CQUIN scheme in 2010/11 had better outcomes on that topic in 2010/11.

The selection of goals in local CQUIN schemes is endogenous and, if not accounted for, is likely to lead to biased estimates of the effect of goal inclusion on outcomes. We assume that this endogeneity is related to previous performance, because the commissioner and provider agree there is weak performance in this area and a need for improvement. We estimated difference-in-differences models with and without hospital fixed effects. The estimated impact of inclusion of a particular goal in local CQUIN schemes therefore indicates whether the *changes* in outcomes in these Trusts differ from the *changes* in outcomes observed for other Trusts. The validity of this approach assumes that decisions over whether to include a topic in a local CQUIN scheme is not systematically related to the anticipated change in the absence of this intervention.

We use indicators published on the NHS Indicators Portal, available before and for the 2010/11 financial year, and related to goals included in a reasonable proportion of local CQUIN schemes in 2010/11. We analyse nine indicators in total, covering patient safety, elective surgery, emergency readmissions, and returns to usual place of residence.

## **Analysis of the structure and content of the 2010/11 schemes**

Analysis of the 2010/11 CQUIN schemes illustrated the wide variation in schemes, topics and indicators across providers, raising questions about the extent to which national monitoring of impact with such local variation would be possible.

### **Descriptive analysis**

The national data set contained 337 CQUIN schemes. The schemes apply to 151 acute care providers, nine ambulance service providers, 93 community care providers and 84 mental health and learning disability service providers.

Overall, the goals in the schemes covered 113 distinct topics. Some of these topics occurred in more than one sector. Overall, for acute care there were 92 distinct topics. For ambulance care there were 29 distinct topics, for community care 63 distinct topics and for mental health 57 distinct topics.

Across all of the 337 schemes, the goals were operationalised using a total of 5,001 indicators. Of these, 3,142 indicators were uniquely-defined indicators. There were 1,546 distinct indicators used in CQUIN schemes for acute care. In ambulance service, community care and mental health schemes respectively, the numbers of distinct indicators were 78, 999 and 645.

Only acute care schemes contained national indicators. These constituted 12% of the total number of indicators in acute care schemes. Fifty-seven percent of the indicators in the acute care schemes were locally developed, and the remaining 31% were regional indicators. In ambulance services the majority (90%) of indicators were regional, while the majorities of indicators in community care (82%) and mental health (64%) were local.

The local flexibility in the design of CQUIN schemes led to schemes that were often highly complex. Table 1 displays CQUIN scheme size and complexity as measured by the number of indicators per scheme in each of the four sectors. For acute care, a single CQUIN scheme could have up to 25 different goals with the median scheme containing 11 different goals. These goals were operationalised with up to 52 different indicators per scheme, with the median scheme containing 16 different indicators. In general, the schemes were less complex in other sectors. Using the number of indicators as a measuring rod for scheme complexity, mental health CQUIN schemes were the second most complex schemes, followed by community care and then ambulance care schemes.

### **Identification and operationalisation of local needs for quality improvement**

Table 2 shows that the most frequently included goal was patient/user satisfaction and involvement. In addition to the nationally mandated patient experience goal included in all acute care schemes, 37% of local goals in acute care were concerned with patient/user

satisfaction or involvement. This goal was included in 89% of schemes in ambulance care, 77% of schemes in community care and 86% of schemes in mental health. The greatest commonality in goals across sectors was between acute and community care, where seven of the 10 most frequently occurring goals in each sector are used in both sectors. Ambulance services, and especially mental health, have more sector specific goals amongst their top ten goals.

In each sector, the least frequently applied goals in the sector are included in at least 22% of the sector's schemes, and on average, goals on a sector's top ten list occur in about 40% of the sector's schemes. This suggests a relatively high degree of agreement within a sector on which local goals require attention. In addition, this degree of agreement between schemes suggests that providers could potentially be benchmarked against a solid base of peers with similar interest, if similar indicators had been used across schemes. However, the Department of Health's hope that locally negotiated schemes would rely on standard performance indicators where possible, and only develop local indicators when no suitable standard/national indicator was available, does not seem to have been fulfilled. Indeed, of the 5,000 indicators in use in the 2010/11 CQUIN schemes more than 60% of these indicators were unique.

As evident from the previous analysis, 'discharge planning and communication' was a popular goal included in about half (72) of the CQUIN schemes for acute care, in 22 community service schemes and in 10 mental health schemes. To operationalise the goal, the 74 acute care provider CQUIN schemes made use of 114 indicators, of which 82 were unique. The median number of indicators used to operationalise the discharge planning goal in local acute care schemes was three, with a minimum of one and a maximum of eight indicators used.

Table 3 lists some of the indicators that were used to operationalise this goal. We only show the local indicators that appeared in more than one acute care schemes. The list of indicators reveals that several of the indicators are concerned with very similar aspects of discharge planning, but the freedom to develop local indicators meant that this commonality of interest could not lead to benchmarking exercises across providers, because very few schemes include the same indicators. For example, many of the indicators consider the information sent to patients' GPs when patients are discharged. While the indicator in one scheme requires discharge summaries to be received by the patient's GP within 24 hours, another scheme requires discharge letters to be received by the patient's GP within 2 weeks of discharge, and yet another requires inpatient and outpatient letters to be received by the GP within 1 week.

A potential solution for the costly and sometimes delayed process of developing indicators was suggested in the CQUIN guidance. The guidance advised commissioners and providers to use official indicators where possible, and only use or develop local indicators when no national alternative was available. The guidance contains a link to the NHS Information Centre website that lists already developed indicators to measure quality. This list currently includes at least 18 indicators related to discharge planning.

Although it is not possible to find national alternatives to all of the local indicators, it is obvious that several of the locally developed indicators could have been substituted by a national indicator. For example, several of the locally developed indicators are concerned with patients' experience of the discharge or the level of information given to patients. In these cases there are obvious national indicator alternatives e.g. PE17-19 and PE25 in the Measuring for Quality improvement (MQI) list. Similarly, schemes choosing to focus on the level of readmissions could have used the national indicators for the areas where these are defined.

### **Compliance with guidelines for scheme design**

Table 4 shows that, although CQUIN schemes were intended to generate changes in outcome, in all sectors, the majority of indicators were classified (by scheme developers) as process or structure indicators. As stated in the CQUIN guidance, process and structural indicators may still be justified provided they link clearly to outcomes. Based on scheme developers' own classifications, just 19% of locally selected structure or process indicators were evidence based.

In 2010/11, scheme developers were advised not to reward data collection. However, inspection of the baseline values associated with the locally-selected performance indicators shows that for 45% of the indicators the baseline value was "TBC" (to be confirmed), and for 24% of the indicators "N/A" (not available). While one of the benefits of negotiating schemes locally was the possibility of defining "appropriate" local stretch, this is difficult if new indicators are being developed and the local benchmark is unknown. While this may be seen as an intermediate phase, with no restrictions on the development of new indicators in the future, this limitation may persist in the future.

It can also be seen from Table 4 that in all other sectors than ambulance service, the majority of local indicators are concerned with improving effectiveness. It is also clear from the table that indicators for innovation are the least common.

One of the problems with different definitions and interpretations of classifications for indicators, is that similar indicators might be classified differently in terms of domains, which raises questions about the extent to which asking for indicators to be classified in this way is meaningful or desirable. For example the data shows that indicators measuring processes of care, for example the proportion of stroke patients given a swallow test screening within 24 hours of admission, were sometimes classified as addressing patient safety, effectiveness and patient experience but in other cases not seen as addressing patient experience.

The confusion around the definition of, and reluctance to incentivise, innovation is also reflected in Table 5 which displays the proportion of local schemes that failed to include an indicator from each quality domain and the proportion of schemes that lack an indicator for each of the four domains. In all but ambulance care, more than half of the locally negotiated

schemes failed to fulfil the requirement of containing at least one indicator from each domain.

# Qualitative analysis of the CQUIN process

## Year 1 (2009/10)

Respondents reported that during this period CQUIN was largely a trial run, or a paper exercise.

*“some of the CQUIN requirements were a little vague, so it was quite difficult I think for commissioners to say well actually yes that’s what we wanted and that’s what we asked for and for us to go well actually that was the information you required. So it was kind of a bit of testing I guess really for the next sort of phase for this year”* ID3 Provider

SHA approaches varied, with responses ranging from mandatory regional CQUIN goals, to optional ‘pick lists’ of indicators. The tight timetable left little room for meaningful dialogue and in many cases indicators were chosen which it became apparent were problematic, but this only came to light when commissioners and providers met as part of the in-year monitoring process. For example, in one case reducing ‘average’ length of stay it was not clear whether median or mean was the measure. More generally, many respondents reported an absence of clarity on indicators and/or payment rules, which resulted in providers receiving CQUIN monies even where performance was not in line with goals.

*“the original targets we did were very crude really because neither knew properly about how we would evaluate it and measure it properly”* ID153 Provider

Another problem was that providers agreed to schemes, without sufficient consideration of data collection requirements and capacity.

*“one of the biggest things that we learned from doing the first year was being really clear about having systems to measure the things that you’re intending to measure, because quite often they didn’t have. So they would sign up to things that the systems weren’t in place, so it was difficult for them to achieve”.* ID54 Commissioner

Given all the problems that were experienced in the first year, commissioners tended to pay all of the CQUIN money to providers, even where performance was poor.

*“we had a very poor year last year on the year one of the discharge summary for instance but apart from squirming at the contract meetings I never felt like there was ever going to be a financial penalty”.* ID15 Provider

Various reasons were given for selecting topic areas and goals including poor performance, local/regional/national priority and improving implementation of guidance, toolkits or pathways. Indicator payment weights sometimes reflected effort (more payment for more effort). In other cases indicators were weighted equally and another approach was to give greater weight to high priority goals.

*“if it's harder and a bigger piece of work it gets a higher weighting. It's as simple as that”. ID23, Commissioner*

*“We all felt that the post-stroke one ... well there was more quality to be gained by achieving that one than the others. So we thought that deserved a higher weighting”. ID18, Provider*

*“ourselves and the PCT understood at an early stage which were the more important ones and therefore should have the bigger weighting and which ones shouldn't”. ID19, Provider*

In part, the lack of sophistication regarding payment weights may reflect the timescales for scheme development and agreement.

The tight timetable left little time to engage front line clinicians, with the result that provider representatives entered into agreements without input from these staff.

*“It was very led by financial leads in the first year, senior financial people in the acute, and there was really poor clinical engagement” ID54 Commissioner*

Commissioners and providers were already meeting regularly as part of the quality contractual monitoring process and CQUIN discussions and monitoring were usually incorporated into these meetings.

Interviews undertaken towards the end of this financial year highlighted lessons learned and the need to avoid these mistakes in future years. Greater involvement of front line clinicians, improved clarity and sophistication of goals and a much longer lead in time were seen as essential moving into Year 2 of CQUIN implementation.

## **Year 2 (2010/11)**

Participants were unanimous in their opinion that the increase in the percentage of income linked to CQUIN goals from 0.5% of contract income in Year 1 to 1.5% in Year 2 had been significant in terms of the impact on the process, with CQUIN receiving greater attention at all levels of the organisation as a result.

*“the increased amount of the financial value I think it's just give the added impetus for organisations to really, really focus and concentrate their efforts in improving quality and identify exactly where quality improvements can be made, within the organisation”. ID17, Commissioner*

However, this does not mean that Board's were necessarily having meaningful conversations about quality and innovation. A number of participants described how the financial risk associated with CQUIN was of key importance in focusing Boards, which may

mean that Boards conceptualise CQUIN in terms of money and targets rather than quality and innovation.

*“Oh the board are very aware of CQUINs, absolutely. Yeah. They’re fully briefed..... and they’re fully aware of the risks associated with the income on that... it was unanimously agreed on the board that the biggest risk to our income is actually CQUIN, cos it’s £Xm”.*

ID4, Provider

Fears were expressed by some providers that in the context of financial constraints commissioners might use CQUIN as a means of withholding money from them. Some commissioners were at pains to emphasise to providers that this was not the case, although comments by some commissioners suggested that not everybody in their organisation would be disappointed if providers failed to meet their CQUIN goals.

*“it’s money, their money, they don’t want to give it to us. And there is a real kind of a sense of feeling about that, well they don’t want us to achieve the CQUIN anyway because obviously they’re going to have to cough [X] million”* ID22, Provider

*“I was talking to the director of finance who was saying ‘if they don’t get that then great, because they won’t get the money’ and I said ‘well I know but equally they fail on cancer targets and our population don’t get the quality of care’”.* ID16, Commissioner

There was no opposition in principle to national goals and they were seen by some people as preferable to local goals, due to the ability to use these for benchmarking and/or specific local goals had proved to be problematic or viewed as inappropriate.

Greater effort had been expended on defining goals and agreeing in year monitoring measures and processes in Year 2 although observation of meetings suggests that there were still problems relating to this issue. In many cases, in reaction to the problems identified in Year 1 relating to poorly specified goals, there was a tendency to describe goals and payment rules in great detail.

Some participants suggested that CQUIN was not an appropriate vehicle for incentivising innovation due to its link to performance within fixed timescales and measures and the constraints of non-recurrent funding. Efforts to tightly specify CQUIN goals and increase in-year performance measures were perceived by some as likely to exacerbate this problem.

*“the real clinical innovation stuff, yeah, has tended to be joint projects that we’re working on anyway... the trouble with CQUIN is that it’s within that year, and things often don’t fit into that timescale.... a lot of the innovation, they’re happy to get on and do, and why would you then add something on that adds a ... bureaucratic system, when actually if you just leave them to get on with it they probably would have delivered it anyway and in a better timescale?”* ID34, Commissioner

*“I think that the template is nice though it is assumes a particular kind of indicator which is not always appropriate and not always useful. So the template almost stops*

*you from doing the things that you would like to do and all that kind of stuff". ID20, Provider*

In many cases, in Year 2, CQUIN schemes were not agreed prior to the contracting year. Many providers did not receive their total CQUIN payment for Year 2 due to underperformance on CQUIN goals.

Interviews undertaken towards the end of this financial year highlighted lessons learned and the need to avoid these mistakes in future years. Greater involvement of front line clinicians, improved clarity and sophistication of goals and a much longer lead in time were seen as essential moving into Year 2 of CQUIN implementation.

The view that CQUIN was improving quality was a relatively common one amongst CQUIN leads and whilst many participants talked about this in general terms, many also gave concrete examples of quality improvements arising from CQUIN goals. However, engagement of front line staff was still a problem with timescales making this difficult and by and large, no mechanisms developed to do this in a meaningful way.

*"the local commissioners meet regularly with non-clinical managers and it's very difficult for clinicians to free up time to get involved in that" ID131 Provider*

Approaches varied in relation to scheme content, with some commissioners seeking to impose goals and others engaging in collaborative dialogue to reach agreement.

In Year 2, although the CQUIN guidance was amended to stop payment to acute providers for collecting data, and the emphasis placed on payment for improved performance, commissioners and providers reported continuing to pay for data collection and viewed the guidance as unhelpful. (This is also evident from our national picture of CQUIN schemes).

*"The strict definition or the strict interpretation of the CQUINs guidance seems to suggest that you can't pay for data collection .... but I think that you have to do that especially if you are asking for new data collection and our providers were willing to do that where they were guaranteed the payments of the first quarter on submission so they were guaranteed to get at least the income for the data collection". ID24, Commissioner*

### **Year 3 (2011/12)**

CQUIN guidance was amended to permit CQUIN to reward data collection. Commissioners and providers welcomed this development and many commissioners reported that they were incentivising data collection in 2010/11 despite the guidance. This was seen as a necessary first step to quality improvement in areas where baseline data were not routinely available.

Local development of indicators was a feature in many sites.

*“probably sixty percent of the CQUINs they said, ‘We want a CQUIN in this area’ but then we had to go away and develop it. So they choose the topic or the area, but we still had to develop the detail behind it...”*ID116 Provider

In many cases providers were not meeting CQUIN goals and monies were withheld accordingly. However, whilst some commissioners applied a strict interpretation of the agreed CQUIN scheme payment rules, others were more willing to take into account mitigating circumstances.

Interviews with front line staff highlighted their lack of involvement with the process, scepticism about benefits and a belief that data collection detracted from time spent with patients.

*“...certainly the feedback that I’ve had from the negotiations for these ones, the 11 to 12 ones from the clinicians, is they were a little bit hurt that they are obliged to, even if it’s something they were doing anyway, that effectively the whole process gets swept up and then they get told what they have to do and what they have to report on and that they haven’t had a chance to influence that.* ID124 Provider

Managers recognised that more effort was needed to involve clinicians and saw this as a priority. However, despite this recurring theme year on year, little progress appeared to have been made in this respect.

*“...I have to say I think we anticipated that this year would run a lot smoother, however, because we’re dealing with the clusters and the PBC now and the GPs are largely dictating what our indicators are it’s not gone to time. I think had it have been the previous years, we’d have been all signed off ready to go and we’d be flying by now [early August 2011], but because we had a big event in February where the GP leads were there and which went well, we understood it from their perspective, but then we’ve had cancellations of meetings because the GP leads aren’t available, but then again we’ve had meetings scheduled in for ages...where they’re supposed to be there and they’re not...”* CQUINID130 Provider

*“They....expect to be able to pass instructions down....that feeds down and somehow at the bottom of the pile the magic happens ...I can tell you it’s not been generated from the bottom up. It’s not even been generated from the middle up, it’s next to top...”* ID135 Provider

*“It was all a done deal before we were even included and that’s caused lots of problems because it’s been back and forth saying, “well this is unrealistic and that’s not how you do this.” So yeah, if we’d been involved in the negotiation from the outset it probably would have been a lot smoother. We absolutely agree in principle with the aims but the implementation process has not been great.”* ID111 Provider

## **Changes in Clinical Behaviours/Practices to improve Quality of Care**

Observational and interview data suggest that CQUIN has made some impact on clinical practice and service delivery. However, reservations were expressed by some respondents about the extent to which CQUIN was improving the quality of patient care. In some cases, maintaining changes in practice was reported as difficult, with constant efforts being required to ensure that clinicians did not revert to pre-CQUIN behaviours and practices. This suggests that some of these changes were not becoming embedded. Ideally those charged with implementing CQUIN at local level want changes which become institutionalised, since this reduces the time and effort associated with monitoring compliance. Additionally, given the potential for removal of incentives in an area to result in a decline in performance levels, practices which have become part of the 'taken for granted' activities of the provider may fare better compared with those which have not, following changes to CQUIN schemes which prioritise new areas.

Although CQUIN is intended to incentivise innovation, participants were unclear about what constituted innovation. Linking performance to financial incentives was seen by some as inhibiting innovative approaches, which might pose a financial risk.

### **Impact of CQUIN implementation on workload**

Evidence suggests that CQUIN resulted in increased levels of workload for all providers. A huge amount of time and effort went into data collection to demonstrate performance. In particular, the data collection associated with monitoring and reporting was reported as placing a heavy burden on providers.

For mental health and community providers the fact that data collection and reporting systems in some areas were embryonic or absent exacerbated the problem. However, also an issue for acute providers where CQUIN goals related to topics which were not covered by existing systems.

### **Impact on relationships between stakeholders**

For some organisations the relationship between lead commissioner and providers was reported as having improved over the period since CQUIN was introduced. In some cases CQUIN has been cited as a key factor in the strengthening of these relationships. Although increased contact between commissioners and providers added to workload (as discussed above), it was also reported as enhancing mutual understanding and broadening discussions to include clinical input.

In around half of the case study sites, most of the Quality Leads at the commissioning bodies and within provider organisations had worked together previously and had developed an understanding of one another's organisations and ways of working. They

already had established quality monitoring processes and therefore built on these existing processes and forums to address CQUIN framework as well.

In a small number of case study sites the relationship has been generally contentious and less friendly; however, CQUIN does not seem to have played a significant role in deterioration of the relationships. There are a small number of sites where CQUIN has been reported as impacting negatively on the relationship between the commissioner and provider organisations.

### **Summary of the CQUIN process**

In general, CQUIN has not impacted hugely on relationships between stakeholders and the impact on quality appears to have been disappointing. In terms of the theory and mechanisms contained in the IA, this does not appear to be related to the size of the incentive as the increase to 1.5% was reported as securing Board attention across all organisations. However, conceptualising CQUIN as posing a financial risk, does not necessarily translate into meaningful conversations about quality and innovation.

Negotiating and developing CQUIN schemes combined with data collection and monitoring involve a huge amount of effort and it is not clear that this is justified, in terms of the benefits of CQUIN. The fact that CQUIN has been implemented in a turbulent policy environment has impacted negatively, although there are aspects of the design and implementation of the CQUIN framework which also appear to detract from its impact. These include changing CQUIN goals and schemes annually and related to this, insufficient mechanisms and time to engage clinicians. Additionally, the local nature of goals makes benchmarking, which can act as a spur to motivate providers, difficult, if not impossible. Furthermore, the fact that many goals concern processes as opposed to outcomes and are based on at best, weak evidence of effectiveness, may also contribute to limited clinical engagement. These features contrast with other financial incentive initiatives (such as Best Practice Tariffs and Advancing Quality) which appear to have greater buy-in and impact.

## Quantitative analysis of the impact of CQUIN goals

Table 6 provides the definitions of the nine indicators that we analysed and how frequently the corresponding topic was included in the local CQUIN schemes in 2010/11.

The regression results for patient safety are shown in Table 7. For both safety incidents and severe safety incidents, the inclusion of patient safety or risk assessment as a topic was associated with a lower number of incidents prior to the introduction of the 2010/11 CQUIN, significantly so for severe incidents. The difference-in-difference estimates are not significant and thus inclusion of the goal was not significantly associated with a *change* in the number of incidents.

Table 8 shows the results for average health gains from elective surgeries from the PROMs dataset. Again, the difference-in-difference estimates of the effect of including surgery as a topic in CQUIN schemes are not statistically significant and the inclusion of surgery as a goal was not associated with a change in average health gains for the PROM measures.

The impact results for rates of return to usual place of residence within 56 days of admission for stroke and hip fracture and emergency readmission within 28 days are shown in Table 9. Inclusion of stroke or hip fracture as a goal in the Trust's CQUIN scheme was not associated with the return rate to usual place of residence after stroke, but the fixed effect difference-in-differences estimate did show a significant increase in the return rate after hip fracture for trusts including a hip fracture as a CQUIN goal. Inclusion of discharge planning in trusts' CQUIN schemes was also not associated with a change in the readmission rate.

Overall, we find no evidence of an impact of the content of local CQUIN schemes on performance improvement except for hip fracture. None of the other estimated impacts is statistically significant and the insignificant estimated coefficients represent variously negative and positive effects on quality.

## Conclusion from CQUIN evaluation

Assessing the impact of CQUIN on quality is problematic due to the wide range of schemes and indicators which characterise the national CQUIN picture. Based on the evidence in our report, however, the impact has been disappointing. Whilst the theory underpinning CQUIN has some validity, its implementation departs substantially from that theory. Financial incentive schemes can place a heavy burden on participants. It is important, therefore, that we learn lessons from CQUIN and other schemes if the benefits of such schemes are to outweigh the costs.

The CQUIN framework was intended to allow providers and commissioners to agree on P4P schemes locally, to ensure that performance goals and indicators included in the scheme would address local needs for quality improvement and generate local enthusiasm through local clinical involvement in the development of indicators. This process was anticipated to lead to quality improvements across the NHS, and over time.

We find that the framework has been successful in helping commissioners and providers to identify and prioritise local needs for quality improvement. However, our research has identified a number of aspects of the current implementation of the scheme that can potentially impede the ability of the scheme to generate the desired improvements in the quality of care.

Firstly, although schemes were developed locally, clinicians expressed dissatisfaction with the scheme development process which often did not involve frontline clinicians, and was acknowledged by managers and clinicians to be deficient in this area. This coupled with the absence of mechanisms for meaningful engagement suggests that the schemes are unlikely to lead to the anticipated local enthusiasm.

The impact assessment published alongside the CQUIN Framework anticipated that standardisation of indicators across schemes would emerge. However, based on the second round of schemes it seems unlikely that this will occur without firmer regulation in place. For example, although the CQUIN guidance stipulated that all schemes include an indicator for each of the quality domains safety, effectiveness, patient experience and innovation, the majority of schemes failed to meet this requirement. The variation in the classification of identical indicators across schemes questions whether letting scheme developers perform this classification is desirable. Also, although the guidance emphasised a focus on outcome, many locally agreed indicators concern structure and processes and are based on, at best, weak evidence of effectiveness. This may also contribute to limited clinical engagement.

Furthermore, the indicators that emerged from the local development process were often unclear or lacking precision which led to difficulty in the follow up of performance. The reliance on locally developed indicators also hinders benchmarking across providers, even for schemes with common goals for which national indicators did exist.

Freedom to use local indicators chimes with the national policy direction which emphasises local freedoms and priorities in public services. Furthermore, in a context where policy makers are keen to engage clinicians in commissioning health services, emphasising local freedom and discretion is more likely to achieve this aim, compared with restrictive approaches that mandate sets of indicators. However, as our findings illustrate, the development of local indicators as part of the CQUIN process has been somewhat problematic, leading to a muted potential impact of the policy.

Designers of P4P schemes should consider two major design aspects: (i) the identification of objectives that need performance improvements; and (ii) the technical design of a P4P scheme that can lead to the desired changes in performance. A regulator wishing to incentivise better quality in lower hierarchical levels of an organisation can in principle choose between top-down performance objectives and technical scheme design or a bottom up approach in which lower hierarchical levels play a role in the design of schemes. Due to asymmetric information between the top and bottom levels of the organisation, it may be sensible to allow considerable local discretion in the choice of goals but, to ensure incentive compatibility, the technical design of P4P is best managed in a top-down fashion.

We have found that the bottom up approach inherent in the CQUIN framework has worked well in terms of identifying relevant goals for inclusion in P4P scheme, whereas the technical design of the schemes was not consistent with the design requirements set out by the DH in an attempt to manage the design process. This may explain why the inclusion of particular goals in local CQUIN schemes did not lead to statistically significant performance improvements in related outcome indicators.

While there is clearly an important case for local strategic and clinical input into the design of pay-for-performance schemes, this should be separate from the technical design process, which involves defining indicators, agreeing thresholds, and setting prices. Defining good performance indicators requires evidence based knowledge of the relationship between structures, processes and outcomes, and providers' ability to affect these measures. Agreeing thresholds requires a baseline and understanding how sensitive performance is to prices. Setting prices involves knowing the value of the desired performance improvements to the commissioner and society (essentially their willingness to pay) and the costs to the providers of making the improvements. Previous research only provides limited knowledge about how this technical process should be carried out. It is thus unlikely that such expertise can be expected in local health economies. The CQUIN framework potentially offered an opportunity to learn how technical design influenced outcome, but due to the high degree of local experimentation and little systematic collection of key variables, it is difficult to derive lessons from this unstructured experiment about the impact and importance of different technical design factors on the effectiveness of pay-for-performance.

Balancing policy goals of localism with the objective of improving patient outcomes leads us to conclude that a firmer national framework would be preferable. This might take the form, for example, of a 'pick list' of national indicators from which commissioners and providers can choose a subset to fit their current priorities. Our interviewees did not reveal major opposition to national indicators and this could be combined with standardised

reporting on the local schemes. This standardised reporting could include *ex post* reporting of the results and the levels of payment as well as *ex ante* reporting of the design of the schemes. These would allow systematic learning across commissioners and providers on how to design a local pay-for-performance scheme, increasing the likelihood that the goals set out in High Quality Care for All could be achieved in the future.

# Findings from our evaluation of other relevant financial incentive initiatives

## Introduction

2 recent initiatives introduced in the NHS in England differ markedly from CQUIN in relation to design and implementation. We outline relevant features of these below to enable us to draw lessons from the evaluation of all 3 schemes in the final section of the report.

## Advancing Quality and Best Practice Tariffs

Advancing Quality (AQ) is an initiative that aims to improve patient care and patient experience in NHS hospitals across 1 region (the North West) of England. It is a voluntary programme which provides financial incentives for improvement in the quality of care provided to patients. The programme is based closely on a project implemented in the US, called the Hospital Quality Incentive Demonstration (HQID), involving a partnership between the Centers for Medicare and Medicaid Services (CMS) and Premier Inc., a nationwide organisation of not-for-profit hospitals. AQ was initially designed and supported by Premier Inc., involved similar quality indicators and financial incentive structures.

The financial incentive programme covered 5 clinical areas as follows: acute myocardial infarction (AMI), heart failure, coronary artery bypass graft (CABG), pneumonia, and hip and knee replacement.

As we outlined earlier, our evaluation of AQ found that it was associated with a clinically significant reduction in mortality for pneumonia patients<sup>3</sup>.

The Best Practice Tariffs (BPT) initiative created incentives to improve care by paying a higher tariff for episodes of care which were in accordance with 'Best Practice' as defined by the BPT scheme. BPTs were initially in 4 areas – cholecystectomy, stroke, hip fracture and cataracts. Our evaluation found that incentives were associated with improvements in processes and outcomes for some procedures<sup>8</sup>.

## Common features of AQ and BPT

*Performance measures are stable over time*

CQUIN schemes involve introducing new indicators and drop existing ones on a regular basis. This can result in a reduction in performance in areas where indicators are dropped from the incentive scheme. In contrast, for BPT and AQ there was no intention to change indicators on an annual basis. This meant that participating organisations could take a longer term approach, investing time and resources to change pathways, employ additional staff where necessary and so on.

### *Measures apply to all participating organisations*

A common measurement set for all participating organisations meant that the AQ programme was able to tap into the competitive spirit of participating providers so that feedback on performance relative to other organisations acted as a spur to improvement. Similarly, for BPT, some clinicians welcomed the ability to use national performance data to drive improvement locally.

### *Focus on a small number of indicators linked to high impact changes*

AQ and BPT which have been shown to have positive impacts on quality are underpinned by this principle, which appears to facilitate a more focused and sustainable approach to change than some other schemes. In contrast CQUIN schemes often contain large numbers of indicators, which combined with the annual process of retiring indicators and introducing new ones, may make sustained effort and institutionalisation of behaviour change difficult.

### *Fit with other levers and policies*

AQ and BPT were implemented in a context characterised by a range of policies and activities aimed at improving care. However, their application was also intended to take account of local contexts and national policies. For example, the HQID scheme on which AQ was based required that pneumonia patients receive antibiotics within 4 hours of arrival. But, in recognition of a national NHS 4 hour wait target for Emergency Department (ED) patients, AQ used 6 hours. Whilst admitting patients a few minutes before they breached the 4 hour ED target was not ideal, rather than trying to change this practice, the leaders reasoned that 6 hours would be long enough to enable hospital staff to cope with a last minute admission to meet the 4 hour target and give 2 hours thereafter to administer antibiotics. Taking account of other constraints in the system to produce measures that would be regarded as reasonable and feasible was seen as important in securing initial commitment and sustained participation therefore. This was made easier by the relatively long (compared to CQUIN) lead in time which allowed for consultation and adaptation to 'real world' constraints.

### *Clinician input*

Encouraging clinical teams to develop local priorities and related incentivised measures might be expected to improve engagement and such thinking certainly informed the development of the national CQUIN scheme. However, indicator development is a time consuming process, which may also lead to a proliferation of local indicators. A further problem is that where clinical teams seek to improve quality in a specific area, they may be tempted (understandably) to set easily achievable (rather than ambitious) goals since failure to meet these goals will impact on organisational income. In other words, this may inhibit, rather than incentivise quality improvement. Clinician input into BPT and AQ did not involve local clinical teams developing indicators in a 'bottom up' way. Instead, structures for consultation and feedback were developed at a 'programme' level so that clinicians from participating organisations could contribute to development of the incentive programmes.

### *Clinical engagement*

Getting 'buy in' from clinicians for incentives can be a protracted and problematic process. For some aspects of the AQ and BPT initiatives, objections were raised about the underlying assumptions and related measures. Feedback in these areas led to changes in response to concerns. Even where support existed for the principles of these policies, clinicians reported achievement as challenging, so that ongoing work was required to sustain improvement beyond the initial phase of implementation. The rotation of junior medical staff was reported as contributing to these challenges.

Engaging clinicians and sustaining their commitment is an ongoing process, though this is likely to be easier where there is support for the programme's measures, where a smaller number of measures are involved (to avoid cognitive overload) and where measures do not change annually.

### *No 'all or nothing' payment rules (& no 'gateways')*

Some incentive initiatives require participants to achieve a threshold level in order to qualify for payment. Furthermore participants may have to pass through 'gateways' and failure to do will result in incentive payment being denied, even where targets have been fulfilled. For example in order to qualify for CQUIN payments from April 2013, 'compliance with high impact innovations' will become a prequalification requirement for CQUIN. The danger with gateways and 'all or nothing' approaches to payment is that where organisations fail to meet gateway requirements or anticipate that threshold achievements cannot be attained, they may have little motivation to continue to pursue quality targets. BPT and AQ do not incorporate these types of payment rules and do not create such perverse incentives therefore.

### *Money as a lever*

For AQ and BPT additional resources were invested to deliver service improvements. For example, the additional (anticipated) income from delivering BPT for hip fracture patients was used to support the case for additional staff (orthogeriatricians), which enabled organisations to deliver care in accordance with best practice. Similarly additional nurses were recruited at many AQ sites as part of the process of pathway redesign involved in delivering more 'joined up' care. The 'one off' nature of CQUIN payments, in comparison, meant that providers were reluctant to invest in extra staff for what was a short term change.

### *Public/wider reporting as a lever*

AQ involved a commitment to publicly reporting the results of the programme from the outset. This acted as a spur to improvement and appeared to contribute to sustained participation in the programme. No such arrangements exist for BPTs. However, the National Hip Fracture Database (NHFD) data show relative performance of organisations on

a range of measures, including those comprised in the hip fracture BPT. Furthermore, what one surgeon referred to as being 'in the spotlight from the Department of Health', due to the availability of these data, was also seen as helpful in getting the attention of senior hospital staff.

#### *Data collection is burdensome*

Incentive initiatives usually require participating organisations to demonstrate achievement using data to substantiate this. Data collection can also be an important part of embedding changes to practice. For AQ and BPT participating organisations experienced problems with data collection and saw this as adding to workload. Where data collection was compatible with and/or built on existing processes, clinicians were less likely to see these requirements as overly burdensome. For example, the fact that data collection requirements for the hip fracture BPT were linked to the NHFD was welcomed by clinicians as working with existing systems. Additionally, clinicians reported using NHFD data to benchmark performance and saw data as useful in informing practice, rather than a 'paper exercise' with little use other than to secure payment. In contrast, by far the most complaints in relation to BPT data collection concerned cataracts, where there were no existing systems to collect the data required. In several cases the burden of data collection was reported as preventing providers and/or commissioners from engaging in BPTs in this area. These data were not seen as useful by clinicians and many disputed the underlying premise of the BPT in cataracts.

#### *Implementation in a turbulent environment*

In the context of turbulence, particularly in relation to the commissioning landscape, initiatives which rely heavily on prolonged and in-depth commissioner input may be less effective than those which require less commissioner input. BPTs use national indicators and payment rules and Advancing Quality (prior to CQUIN) used regional rules and indicators. In both cases, commissioners receive feedback on performance, but their input is relatively minor, which means that implementation is not threatened by changes in organisational structures and personnel in commissioning function.

## **Lessons from CQUIN, AQ and BPT for the design of financial incentive initiatives**

### **1. Schemes and indicators should adopt a longer term perspective than the current annual CQUIN cycle**

The requirement to achieve 'stretch' goals, led to much time and effort being expended on developing and agreeing new CQUIN goals each year, with the result that in many cases behaviour change was only temporary and compliance was of a 'tick box' nature<sup>9</sup>. A much longer term perspective is likely to be required if changes are to become institutionalised.

### **2. Where possible, local indicator development should be avoided**

In many cases problems arose because local managers attempted to develop indicators in response to perceived problems. Indicator development is time consuming and ideally involves piloting, rather than a 'big bang' approach. In most sites, expertise in indicator development was lacking. Selecting from a 'pick-list' of validated indicators would avoid the time and effort of developing indicators at each site. Some SHAs circulated pick-lists of indicators, and there are a range of sources in the UK and beyond which contain indicators, but local development of indicators was a feature in all sites. This may be because pick-lists were insufficiently broad. However, where indicators are lacking, it would be preferable to have a more coordinated response to this, which would enable the application of specific expertise and maximise the benefits from this process.

### **3. A small number of indicators linked to high impact changes are preferable to a large number of indicators covering a wide range of conditions**

CQUIN schemes often contained large numbers of indicators which, combined with the annual process of retiring indicators and introducing new ones, made sustained effort and institutionalisation of behaviour change difficult. Initiatives such as Advancing Quality in the NHS North West and BPT which have been shown to have positive impacts on quality are underpinned by this principle, which appears to facilitate a more focused and sustainable approach to change.

### **4. Mechanisms for engaging clinicians should be clearly identified**

A key problem with CQUIN implementation is the lack of clinical engagement. This is exacerbated by tight timescales, but even if a longer term perspective is taken, there is a danger that engagement will be limited unless processes to engage clinicians are clearly articulated. As part of this process, what constitutes 'engagement' (in particular the extent to which clinicians can input into the process, as opposed to being informed about the likely contents of CQUIN schemes) must be defined in order to ensure that any processes are fit for purpose and that expectations of all parties are clear at the outset.

**5. Indicators and reward structures should be designed and used in a way which complements other incentives and levers in the system**

Often staff involved in negotiating CQUIN schemes had little or no knowledge of other financial levers such as BPTs. This may be because of the ways in which communication flows to commissioners and providers, which results in a lack of joined up thinking about the whole range of levers available. As initiatives are discussed and launched, it is essential therefore to make connections with the contents of commissioner toolkits, as opposed to seeing them as a “stand-alone” initiative.

**6. ‘All or nothing’ payment rules should be avoided**

Where payment was conditional on achievement of absolute levels of performance, this was reported as having a demotivating effect in cases where progress was made but absolute targets not achieved. Initiatives such as BPT, which rewards providers for each patient whose treatment is in accordance with Best Practice rules, ensure that providers are motivated to continue all year round. In contrast, where in-year data suggest that absolute target performance levels are not achievable, there is a danger that attention and effort will be diverted elsewhere to other, more achievable goals.

**7. Careful thought needs to be given to weighting of rewards**

Payments to general medical practices in relation to the Quality and Outcomes Framework recognise that some goals will take more time and effort to achieve than others. Payments are weighted to reflect this therefore. We found a range of rationales and weighting rules, many of which were developed in an ad hoc manner. There was often little or no reflection on the influence of these approaches in a way that could inform future years’ schemes and weighting. It would be helpful to devote time to clearly articulating underpinning principles and to evaluating the extent to which these have been successful following their application.

**8. Careful thought needs to be given to costs and benefits of data collection, monitoring and feedback**

Data collection was perceived as a huge burden by many participants. Furthermore, clinicians viewed data collection as a distraction from quality improvement, as opposed to contributing to it. The relaxation of CQUIN rules to permit payment for data collection was welcomed by participants who recognised that effort is required to design and implement data collection in new areas.

Our evaluations of BPT and AQ suggest that clinicians are not generally inclined to take on additional data collection. However, factors such as the extent to which participants view the measures as useful and the stability of measures over time, as well as the extent to which data collection builds on existing data systems are all important in influencing attitudes to data collection. Consideration of the costs and benefits of data collection and monitoring should be undertaken in a way which recognises other features of a scheme’s

design and sees them as interrelated therefore, rather than treating this as a narrow 'technical' issue which can be assessed in isolation from other scheme features.

### **9. Designing for turbulence**

CQUIN is intended to improve relationships and dialogue between commissioners and providers, resulting in improved quality of care for patients. In the context of turbulence, particularly in relation to the commissioning landscape, initiatives which rely heavily on prolonged and in-depth commissioner input may be less effective than those with are more 'light touch' and self implementing than CQUIN. For example, BPTs use national indicators and payment rules and Advancing Quality (prior to CQUIN) used regional rules and indicators. In both cases, commissioners receive feedback on performance, but their input is relatively minor, which means that implementation is not threatened by changes in organisational structures and personnel in commissioning functions. Consideration should be given to designs which free up commissioners, avoiding protracted negotiation and monitoring processes and at the same time, enable providers to continue to focus on achieving goals, even amidst major changes in their partner organisations.

### **10. Designing for benchmarking**

The theory underpinning CQUIN emphasised local ownership, which led to a wide range of schemes, topic areas and indicators. Local ownership does not necessarily have to entail lots of people developing their own indicators. The wide range of schemes meant that meaningful comparisons across providers were difficult. This makes it difficult to use feedback on performance, relative to other providers, as a spur to improvement. The AQ programme tapped into the competitive spirit of participating providers and even led to high performers sharing improvement tips with others, so that competition did not damage collaboration<sup>3</sup>.

### **11. Introducing financial risk may inhibit, as opposed to encourage, innovation**

Financial incentives, such as those contained in CQUIN schemes, are likely to encourage participants to 'play safe' and avoid experimentation, with its attendant financial risk. Careful thought needs to be given to alternative approaches to stimulating innovative practice.

### **12. Financial incentive initiatives should build on evidence of 'what works'**

There is a growing evidence base relating to incentive schemes and indicator development<sup>10</sup>. This suggests that key factors to consider when designing schemes are the extent to which the scheme:

- is perceived as fair and legitimate and, linked to this, whether what is being asked is feasible (this is influenced by a range of factors including the validity and reliability of indicators, the extent to which individuals can impact on performance and the

organisational setting provides the capacity to respond to in the desired way, whether the compensation is sufficient to cover the additional costs of compliance).

- is clear and targeted properly
- threatens norms and cherished identities
- is based on data which are perceived as accurate.

Evidence may also come from past experience in the NHS. With regard to CQUIN, our interview data highlight some learning from previous years. However, in order to do this in a more systematic fashion, commissioners would have to report what they incentivised and how, what was achieved and paid out, and how future plans reflect this learning. This would also enable an assessment of some of the costs of CQUIN alongside the benefits.

## Tables

**Table 1: Numbers of goals and indicators in CQUIN schemes by sector**

Sector	Dimension	Mean	St.Dev.	Minimum	Median	Maximum	Number of schemes
Acute care	Indicators	18.4	9.4	3	16	52	151
	Goals	11.3	4.4	2	11	25	
Ambulance care	Indicators	9.6	4.7	5	8	19	9
	Goals	6.8	3.2	4	6	12	
Community care	Indicators	12.4	5.9	3	12	28	93
	Goals	7.4	2.7	2	7	13	
Mental Health	Indicators	12.7	8.7	1	10	37	84
	Goals	7.7	3.8	1	7	18	

**Table 2: Top 10 goals in local schemes by sector**

Goal	Acute		Ambulance		Community		Mental health	
	Prop. of schemes	Prop. of weights						
Patient/user satisfaction/involvement	37%	4%	89%	15%	77%	15%	86%	21%
End of life	39%	3%	44%	6%	58%	10%	0%	0%
Falls	37%	3%	56%	6%	38%	4%	0%	0%
Tissue viability/pressure ulcers	47%	4%	0%	0%	54%	9%	0%	0%
AMI & Stroke	41%	3%	56%	8%	0%	0%	0%	0%
Smoking	39%	3%	0%	0%	39%	4%	0%	0%
Discharge planning/communications	47%	5%	0%	0%	25%	3%	0%	0%
Maternity	37%	3%	0%	0%	27%	3%	0%	0%
HoNOS/PbR	0%	0%	0%	0%	0%	0%	57%	10%
Data quality	0%	0%	22%	2%	31%	9%	0%	0%
Recovery planning	0%	0%	0%	0%	0%	0%	39%	7%
Dementia	0%	0%	0%	0%	0%	0%	38%	4%
Essen climate scale	0%	0%	0%	0%	0%	0%	36%	6%
Long term conditions/care planning	0%	0%	0%	0%	34%	6%	0%	0%
Asthma	0%	0%	33%	3%	0%	0%	0%	0%
Alternate care pathways	0%	0%	33%	17%	0%	0%	0%	0%
Cardiac care	0%	0%	33%	3%	0%	0%	0%	0%
Surgery	32%	2%	0%	0%	0%	0%	0%	0%
Neonatal units	30%	9%	0%	0%	0%	0%	0%	0%
Nutrition	0%	0%	0%	0%	30%	4%	0%	0%
Structured activity	0%	0%	0%	0%	0%	0%	30%	4%
Access to services	0%	0%	0%	0%	0%	0%	29%	2%
Service specifications	0%	0%	0%	0%	0%	0%	29%	8%
Crisis resolution home treatment	0%	0%	0%	0%	0%	0%	24%	2%
Reduction in average length of stay	0%	0%	0%	0%	0%	0%	23%	3%
Staff development/improvement	0%	0%	22%	6%	0%	0%	0%	0%
Safeguarding	0%	0%	22%	4%	0%	0%	0%	0%

**Table 3: Local indicators used to operationalise a discharge planning goal in more than one acute care CQUIN scheme**

Definition of indicator	Frequency
Patients to receive a copy of their electronic discharge summary on day of discharge.	4
Discharge summaries to be received by the patient's GP within 24 hours.	4
Discharge summaries to contain the recommended CRG minimum dataset.	4
Discharge letters to be received by patient's GP within 2 weeks of discharge	4
Agree a solution and timescale for the implementation of an electronic discharge summary	4
Estimated date of discharge discussed within 24hrs of admission.	4
Ready to go – no delays.	4
Discharge of inpatients prescribing.	3
Improved patient safety by implementation of electronic discharge summaries	3
In-patient letters to be received in general practice within 1 week in 2010-11 as per standard NHS contract	2
Accuracy of medicines on discharge	2
Discharge information from A&E and day case surgery to GPs	2
Increase effectiveness of accident and emergency discharge information	2
Increase in nurse and midwifery led discharge	2
Increase the number of patients in NHS provided care who have their discharge managed and led by a nurse or midwife where appropriate.	2
Managed discharges	2

**Table 4: Scheme developers' classifications of local indicators by quality domain and dimension**

Sector	Domain				Dimension			
	Outcome	Process / structure	Action plan	Data collection	Safety	Effectiveness	Patient experience	Innovation
Acute	34%	62%	5%	2%	44%	56%	35%	8%
Ambulance	13%	88%	0%	0%	25%	50%	88%	13%
Community	32%	62%	6%	1%	39%	53%	39%	12%
Mental Health	15%	76%	8%	3%	21%	47%	41%	15%
Total	29%	65%	6%	2%	38%	53%	37%	10%

**Table 5: Local schemes not meeting CQUIN guidance design criteria**

Sector	Schemes not containing one indicator in each domain	Proportion of schemes that lacked an indicator for:			
		Safety	Effectiveness	Patient experience	Innovation
Acute	62%	10%	6%	8%	58%
Ambulance	50%	50%	0%	0%	50%
Community	70%	23%	20%	17%	59%
Mental Health	57%	20%	11%	17%	46%

**Table 6: Definitions of indicators used to estimate the impact of local CQUIN schemes**

<b>Definition of indicator</b>	<b>Corresponding CQUIN goal</b>	<b>Number of CQUIN schemes including this goal</b>
Proportion of patients discharged to their usual place of residence within 56 days of an emergency admission for stroke	AMI & Stroke	62
Proportion of patients discharged to their usual place of residence within 56 days of an emergency admission for hip fracture	Fractured neck of femur	11
Emergency readmission	Discharge planning	74
Patient safety incidents	Patient safety or risk assessment systems	13
Severe patient safety incidents	Patient safety or risk assessment systems	13
Average health gain from elective hip replacement surgery	Surgery	49
Average health gain from elective knee replacement surgery	Surgery	49
Average health gain from elective groin hernia repair surgery	Surgery	49
Average health gain from elective varicose vein surgery	Surgery	49

**Table 7: Impact of inclusion of a patient safety or risk assessment goal on safety incidents**

	Severe safety incidents		Safety incidents	
	DID	DID FE	DID	DID FE
Apr2009-Sep2009	-0.0224*** (-3.50)	-0.0227*** (-4.66)	0.306 (1.08)	0.311 (1.88)
Oct2009-Mar2010	-0.0249*** (-3.91)	-0.0252*** (-4.85)	0.651* (2.36)	0.656*** (3.82)
Apr2010-Sep2010	-0.0188** (-2.96)	-0.0192*** (-3.48)	0.856** (2.61)	0.864** (3.01)
Oct2010-Mar2011	-0.0109 (-1.46)	-0.0112 (-1.61)	0.982*** (3.43)	0.989*** (3.94)
Safety goal	-0.0197*** (-3.88)		-0.351 (-0.93)	
DID	0.00322 (0.39)	0.00361 (0.37)	-0.741 (-1.30)	-0.780 (-1.39)
Constant	0.0614*** (11.22)	0.0606*** (15.15)	5.137*** (23.35)	5.114*** (34.49)
N	722	722	723	723

Reference period is Oct2008-Mar2009. Robust t statistics in parentheses. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

**Table 8: Impact of inclusion of a surgery goal on PROMs gains from elective surgery**

	Hernias		Hips		Knees		Veins	
	DID	DID FE	DID	DID FE	DID	DID FE	DID	DID FE
2010/11	0.000288 (0.08)	0.000600 (0.17)	-0.0104* (-2.12)	-0.0101* (-2.61)	0.00232 (0.45)	0.00215 (0.46)	-0.0111 (-1.59)	-0.0103 (-1.99)
Surgery goal	0.00297 (0.74)		-0.00147 (-0.25)		-0.00201 (-0.27)		-0.00675 (-0.92)	
DID	-0.00650 (-1.12)	-0.00626 (-1.11)	-0.00677 (-0.75)	-0.00597 (-0.98)	-0.00693 (-0.69)	-0.00555 (-0.67)	0.00384 (0.34)	0.00799 (0.88)
Constant	0.0805*** (33.84)	0.0813*** (61.11)	0.404*** (110.65)	0.403*** (266.95)	0.291*** (83.01)	0.290*** (148.84)	0.0971*** (23.37)	0.0937*** (48.29)
N	231	231	235	235	240	240	130	130

Reference period is 2009/10. Robust t statistics in parentheses. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

**Table 9: Impact of inclusion of CQUIN goals on returns and readmissions**

	Hip fracture returns		Stroke returns		Readmissions	
	DID	DID FE	DID	DID FE	DID	DID FE
2007/8	-	-	-	-	-	-
	-	-	-	-	-	-
2008/9	1.873	0.973**	2.602**	2.547***	0.740	0.634**
	(0.84)	(2.67)	(3.28)	(4.74)	(1.36)	(2.74)
2009/10	3.443	2.490***	5.107***	5.020***	1.121*	0.911**
	(1.56)	(3.53)	(7.13)	(9.19)	(2.08)	(3.25)
2010/11	4.369	3.388***	7.187***	7.117***	1.319*	0.923**
	(1.85)	(4.54)	(7.92)	(10.70)	(2.09)	(2.95)
Inclusion of goal	4.504**		0.881		0.308	
	(3.10)		(1.43)		(0.69)	
DID	3.854	4.218**	-0.799	-0.834	0.459	0.528
	(1.32)	(3.24)	(-0.64)	(-0.99)	(0.52)	(1.12)
Constant	41.64***	42.69***	57.25***	57.74***	9.525***	9.847***
	(25.33)	(113.52)	(101.69)	(163.96)	(22.98)	(64.03)
N	569	569	500	500	566	566

Reference period is 2007/8. Robust t statistics in parentheses. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

## References

---

- <sup>1</sup> Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research* 2010;10:247.
- <sup>2</sup> Flodgren G, Eccles M, Shepperd S, Scott A, Parmelli E, Beyer F. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database of Systematic reviews* 2011;CD009255.
- <sup>3</sup> Sutton et al. Reduced Mortality with Hospital Pay for Performance in England, *New England Journal of Medicine* 2012; 367:1821-8.
- <sup>4</sup> Pawson R. and Tilley N. *Realistic Evaluation* London: Sage, 1997.
- <sup>5</sup> [http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/@dh/@en/documents/digitalasset/dh\\_091690.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091690.pdf)
- <sup>6</sup> Blumer, H (1954) "What Is Wrong with Social Theory?" *American Sociological Review* 19:3-10.
- <sup>7</sup> Strauss A and Corbin J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage Publications.
- <sup>8</sup> Full report here <http://www.nottingham.ac.uk/business/news/documents/bpt-dh-report-21nov2012.pdf>
- <sup>9</sup> E.g. McMurrin et al. The Profile of Risky Single Occasion Drinkers Presenting at an Emergency Department. *Journal of Substance Misuse*.
- <sup>10</sup> E.g. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA* 2007;297:740-4; McDonald, R., White, J. and Marmor, T. Paying for Performance in Primary Medical Care: Learning About and Learning From "Success" and "Failure" in England and California. *Journal of Health Policy, Politics and Law* 2009; Vol. 34, No. 5, 747-776; Campbell SM, Lester HE. Developing QOF indicators and the concept of *Qofability*. In: *The Quality and Outcomes Framework: transforming general practice* . Radcliffe Publishing, 2010. p. 16-30.