(version 1.0, 5 September 2016)

Scottish School of Primary Care

Series

Literature Review

What did we learn from 12 years of QOF?

Bruce Guthrie
Professor of Primary Care Medicine
University of Dundee
b.guthrie@dundee.ac.uk

Jason Tang
Research Fellow
University of Dundee

Scottish School of Primary Care
General Practice & Primary Care,
Institute of Health & Wellbeing,
College of MVLS,
University of Glasgow,
1 Horselethill Road,
GLASGOW G12 9LX
Email: info@sspc.ac.uk

SSPC_News
www.sspc.ac.uk

What did we learn from 12 years of the Quality and Outcomes Framework?

Executive summary

This literature review draws on a systematic search for relevant papers which identified 219 articles, with 98 papers examining change over time, practice organisation, and unintended consequences contributing to the review where relevant.

Did QOF achieve its intended goals?

The 'core bargain' of QOF was that the government would get higher quality in return for GPs getting higher pay, and that this would address the crisis of GP failing morale which was leading to recruitment and retention problems. In the short-term (first 2-4 years) the QOF was associated with higher performance in relation to incentivised quality of care, although, the gains mostly represented an acceleration of existing pre-QOF trends towards higher quality rather than a clear step change. Over time, quality has remained high, but hasn't clearly further improved. In the first two years, QOF was associated with significant increases in GP pay and a large improvement in GP morale. Over time, practice income and pay has declined to below 2004/5 levels, and GPs' satisfaction with their work is back to pre-2004 levels.

Unintended consequences

All improvement programmes have unintended consequences. QOF was perceived by professionals (and patients when asked) to lead to a more bureaucratic, less individualised type of care with an increasing biomedical focus. Observation of consultations provides some evidence of this happening, and personal continuity reduced due to patients finding it harder to get appointments with the GP of their choice. Any focus on QOF inevitably risked that care for other (non-QOF) conditions might suffer. Care for other conditions did not get worse after QOF, but the existing rate of improvement in quality slowed down, consistent with some 'crowding out'. Policymakers were particularly concerned about the risk of gaming, but there is no consistent evidence of widespread gaming with the partial exception of selective coding of depression for the highly-contested depression indicators.

Technical failures of implementation

A number of indicators were not fit for purpose, either because their purpose was poorly defined in the first place (for example, the obesity register) or because their design was flawed (for example, indicators with very few patients per practice; the problematic update to diabetes register definitions). Payment system implementation was problematic, with the initial QOF payment system systematically diverting resource from small to large practices, and from the deprived to the affluent. Changes to QOF were made without planning for evaluation to answer policy relevant questions such as the effect of indicator withdrawal.

Conclusion

QOF worked as intended initially, delivering higher quality to the government and increased income to GPs, with an improvement in the shared problem of morale. The alignment of payments to internal motivation was important, as well as initial substantial support for implementation. In the long run, this 'bargain' broke down with static quality and falling income and morale. Unintended consequences and technical failures are inevitable in all improvement programmes, but a key lesson for implementing the new contract is that they can be minimised by careful programme and indicator design, and by involving policymakers, clinicians and technical experts in target setting and piloting. Although Scotland is planning for local autonomy, clusters and localities will need considerable facilitation and technical support to be effective and avoid predictable technical failures. It is unclear what the impact of the withdrawal of QOF incentives will be, and this will require monitoring.

Introduction

In the early 2000s, general practice was in crisis with falling morale, recruitment and retention problems (despite this time being the start of a period of unprecedented growth in healthcare spending). This was the background to the introduction of the new General Medical Services (nGMS) contract, which replaced the existing funding system of allowances, capitation, and limited fee for service and target payments. The three core elements of nGMS were a revised capitation payment (the global sum), a pay for performance system (QOF), and flexibility to negotiate payments for specific additional services (a variety of enhanced services). There were other important elements to nGMS which are often overlooked, and which had significant consequences. These included significant implementation support including investment in IT and IT support; the public reporting of QOF scores (which was an additional incentive beyond financial reward to many practices to do well on QOF); and allowing GPs to opt out of out-of-hours care in return for a reduction in income (which was so attractively priced that almost all GPs immediately opted out). The focus of this paper is the QOF, and its design is briefly outlined in appendix 1.

Literature reviewed

We systematically searched the academic literature for papers relating to the QOF, identifying 219 documents including 59 research papers examining changes in a range of measures over time (including clinical quality, patient experience, GP satisfaction, and related specialist service use and emergency hospital admission), 25 research papers using qualitative methods to explore perceptions of QOF and the impact on practice organisation and work, 14 research papers whose focus was on various unintended consequences of QOF implementation (although some other research papers also do this as part of their analysis), a number of other empirical research papers including 47 examining the association at one time point between quality of care measured by QOF and various other measures of quality, and a range of reviews, editorials and comment pieces. Many of these papers do not directly address the impact of QOF, including for example cross-sectional quantitative studies examining associations between QOF measured quality of care and practice characteristics or other outcomes like hospital admissions. What follows therefore draws on a selection of identified papers.

Did QOF achieve its intended goals?

Although the original contract documentation lists a number of broad aims, the specific goals of QOF were not that well-articulated. Professor Martin Roland argues that the contract in general, and QOF in particular, was essentially an agreement that, in return for GPs delivering improved or consistently high quality of care, the government would substantially increase practice income to address falling morale and a recruitment crisis.[1] This 'bargain' is reflected in the title ("Investing in General Practice") and in the high-level summary of the documentation sent to GPs prior to their vote on whether or not to accept (for example, to "reward practices for delivering clinical and organisational quality, through the evidence-based Quality and Outcomes Framework which is in line with professional practice, and for improving the patient experience" p2).[2] Quality of care, income and morale are therefore the focus of this section.

Impact on the quality of QOF incentivised care

Properly understanding the impact of QOF requires the consistent measure of quality over time both before and after QOF, which limits attention to a subset of QOF indicators most commonly those related to diabetes and cardiovascular care.

Analysis using routine electronic data found there were substantial improvements in the quality of incentivised care for people with diabetes both before and after QOF implementation, and in the context of already improving quality, the impact of QOF implementation was at best small,[3] with the largest impact observed in the first year (a 14% relative increase in a composite measure of diabetes quality compared to what was expected based on pre-QOF trends).[4] For hypertension, there was also a pre-QOF improvement in blood pressure monitoring, intensity of treatment and control, but no impact of QOF implementation on the rate of improvement.[5] Although there is some variation depending on the indicator, a similar pattern of an initial acceleration of the rate of improvement in the first year was observed across many disease domains, with a return to trend or reaching a ceiling subsequently.[6] Analysis using manual record review of quality of care for asthma, diabetes and coronary heart disease found the same substantial improvements in quality of incentivised care pre-QOF. QOF implementation accelerated the rate of improvement of quality of diabetes and asthma care in the year after implementation, but had no effect on the rate of improvement of quality of coronary heart disease care. Quality for all three conditions subsequently plateaued although this is likely to be a ceiling effect.[7]

More substantial changes were observed in some other areas, notably prescription of long-acting reversible contraceptives (LARC) for which an indicator was introduced in April 2008, with a subsequent substantial increase in the use of injectable LARCs and implants.[8] Similarly, the recording of smoking status and smoking cessation advice rapidly increased at the time of QOF implementation, particularly for people with chronic diseases where the incentives were largest.[9, 10]

However, at least some of this observed increase in quality of care for smokers appears to be due to changes in the recording rather than changes in actual practice, with for example little change in the prescription of drugs for smoking cessation over the same period.[9-11]

Impact on other outcomes

There are relatively few studies examining whether QOF implementation was associated with improvements in longer-term outcomes. An analysis of emergency hospital admissions found that a previous rising

trend in ambulatory care sensitive admissions (ACSA) for conditions incentivised by QOF reversed with the introduction of QOF, whereas admission rates for ACSA for conditions not incentivised and for non-ACSA conditions both continued to rise. The benefits in reduced admissions were substantial and continued to grow over time, although it is unclear what drove this given the relatively modest impact of QOF on incentivised processes and intermediate outcomes.[12]

Fleetcroft et al modelled likely mortality benefit based on observed changes in quality of care, and estimated that QOF implementation reduced mortality by 11 lives per 100,000,[13] but consistent with their observation that QOF did not accelerate the existing rate of improvement of hypertension care, Serumaga et al did not find any impact of QOF implementation on cardiovascular events and mortality.[5] More recently, a study examining changes in mortality in the UK compared to other countries did not find any effect of QOF implementation on the existing rate of decline for either QOF incentivised or other conditions.[14]

Impact on inequalities

Although it wasn't an explicit aim of QOF, there is evidence that variation between practices in QOF incentivised quality of care reduced after QOF implementation, including that small practices and those serving more deprived populations 'caught up' with larger and more affluent practices.[15,16] Patient-level analysis is more ambiguous though, with for example evidence that although rates of flu immunisation substantially increased in all groups in Scotland, socioeconomic gradients persisted,[17] and that inequalities between ethnic groups persisted post-QOF (although not all these inequalities favour white patients).[18,19]

Impact on practice income

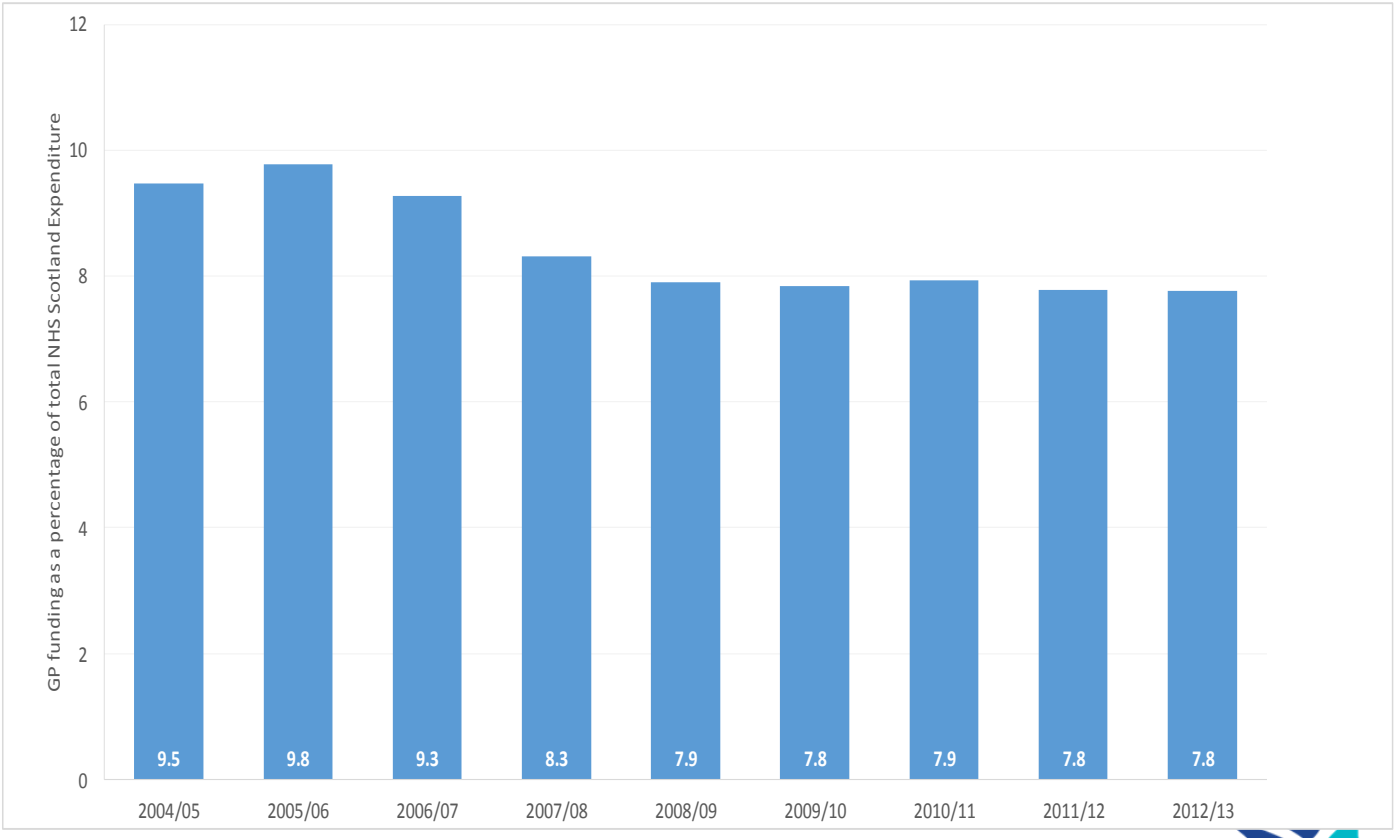
nGMS was intended to deliver a significant increase in practice income in 2004/5 with a further increase in 2005/6. This was associated with an increasing share of NHS spending going to general practice (at a time when total spending was rising at unprecedented rates – figure 1).[20] Over time though, the share of NHS spending going to general practice has fallen (figure 1). In the context of falling total spending, in England this has equated to falling funding for general practice of 1.3% per year in real terms between 2009/10 and 2012/13.[21]

At least some of this increased practice income was invested in increased employment of practice nurses,[22] but GP incomes also rose substantially in the first two years of QOF (2004/5 and 2005/6 – figure 2).[23] Again though, over time GP personal income in real terms has declined to be 11.8% below the level in the first year of nGMS and 18.4% below the peak 'investment in general practice' in 2005/6 (figure 2).

Impact on GP morale

Early comparisons of GP satisfaction with work , comparing 2005 to 2004 found that despite considerable pre-contract apprehension among GPs, there was a significant increase in their job satisfaction, associated with large changes in satisfaction with remuneration

Figure 1: Proportion of NHS Scotland funding going to general practice (adapted from[20])



and hours of work.[24] A longer term view of the same survey shows that overall satisfaction rose in association with the new contract, but has fallen ever since and is now similar to the pre-QOF survey in February 2004.[25] In practice, patterns of overall satisfaction largely mirror patterns for the three individual areas of hours of work, recognition for good work and remuneration all of which are back to 2004 levels.

Figure 2: GP personal income in real-terms compared to 2004/5 (rebased to 100)[23]

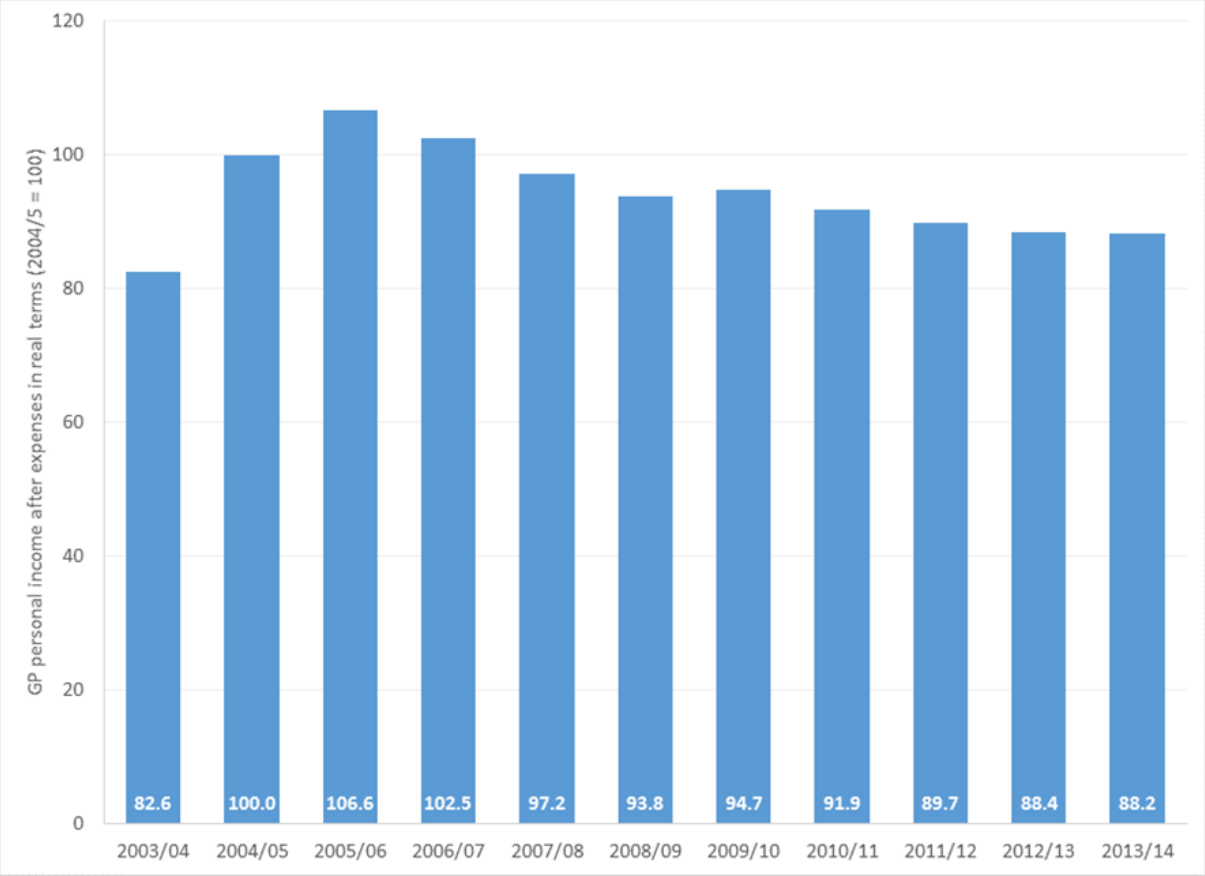
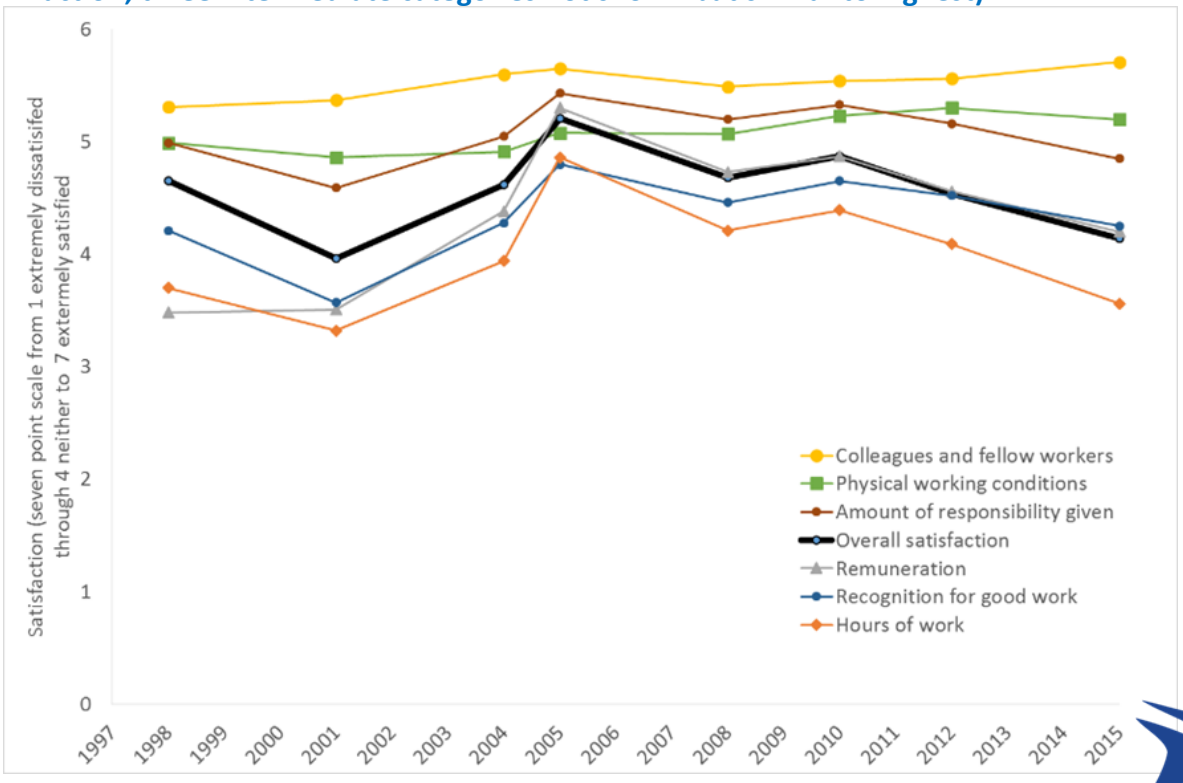


Figure 3: GP work satisfaction in England in the eight National GP Worklife Surveys (overall satisfaction and the three areas with highest satisfaction and lowest satisfaction; three intermediate categories not shown but similar to highest).[25]



Summary

In relation to the 'core bargain' between government and GPs (higher pay to address a crisis of morale and recruitment in return for higher quality), the evidence is that:

- QOF implementation happened at a time when quality of care was already increasing. Overall, QOF implementation was associated with an initial increase in the rate of improvement, but subsequently the rate of improvement returned to trend and/or measured quality reached a ceiling.
- QOF implementation was associated with a substantial increase in the proportion of NHS spending going to general practice and in GP personal income, which over time was clawed back.
- QOF implementation was associated with a substantial improvement in GP morale, which over time has reversed.

From the perspective of the core bargain underlying the new GMS contract, QOF appears to have been an initial qualified success but a long-term failure (although quality of care now remains substantially better than quality of care in the late 1990s and early 2000s).

Negative unintended consequences of QOF implementation

All significant organisational change has negative unintended consequences. Most unintended consequences are, in general, highly predictable (for example, pay-for-performance will predictably risk tunnel vision and gaming) but the specifics of how unintended consequences play out in a particular context are not so predictable, and a particular concern of clinicians in relation to QOF has been that practice has become more bureaucratised ('box ticking') and less personal.[26, 27] Box 1 summarises some of the examples of negative unintended consequences.

Bureaucratisation, clinical autonomy and professionalism

A common feature of editorials and comment features is concern that QOF would lead to a reduction in clinical autonomy by driving 'tick-box' medicine through the application of standardised protocols to all patients irrespective of their circumstances.[28, 29] More consistent delivery of care is of course central to improving quality, and exception reporting was designed to ensure that clinicians could tailor care to individuals without financial penalty to themselves. A number of qualitative studies using interviews and, less commonly, observation found that practices reorganised themselves to ensure delivery of QOF-incentivised care, by creating or more systematically using registers, recall systems, electronic medical records, and chronic disease management clinics.[27, 30-35] These organisational changes were paralleled by changes to inter- and intra-professional relationships. Significant areas of clinical work were de novo or more completely delegated to practice nurses, but administrative staff also took on new responsibilities for recall systems and monitoring professional work to ensure delivery.[36] Additionally, at least for QOF work, there was some evidence of stratification among GPs with a distinction between those involved in QOF organisation ("the chasers") and other GPs subject to surveillance and management ("the chased").[32] One paper framed this approvingly as showing how structural change was associated with improved competence and efficiency.[30] More

Box 1: Examples of negative unintended consequences in QOF

Bureaucratisation, clinical autonomy and professionalism

Qualitative work found that both professionals and patients expressed concern about care becoming a 'box ticking exercise'. Observational studies found evidence of a greater biomedical focus in consultations with less room for the patient voice, but there is little quantitative evidence in this area (tunnel vision). Patient surveys showed no change in overall satisfaction with general practice or with communication by professionals between 2003 and 2007, but significant falls in the patients being able to see their own GP and satisfaction with this, possibly caused by policy focus on rapid access and/or fragmentation of chronic disease care across multiple nurse-led clinics.

Crowding out

Qualitative studies found repeated concern about negative effects on care not incentivised in QOF. Longitudinal quantitative studies found evidence to support this, with more rapid improvements in QOF incentivised care being matched by a slowing in the rate of improvement for non-incentivised care.

Gaming

High-stakes target-setting and financial incentives will always lead to behaviour change ranging from legitimate maximisation of performance by ensuring complete recording of care already given all the way to outright fraud. There is evidence for changes in recording in response to QOF, notably for the smoking indicators. There is no published evidence that fraud happened any more than it already did. Gaming sits in the grey zone between these extremes, and the evidence on it is somewhat inconsistent. For example, some authors found that there was an unexpectedly high number of blood pressure records just below targets as opposed to just above, but others did not. Most convincingly, consistent with qualitative evidence that GPs were more willing to consider gaming where they did not believe incentivised care mattered, there was evidence of a small shift away from the use of QOF-qualifying codes for depression towards symptom codes after the introduction of the highly contested incentives for severity screening at diagnosis of depression.

commonly, a degree of ambivalence from practice staff was reported (in the context of broad support for the importance of QOF incentivised care), with both GPs and nurses expressing concern about a drift to "ticking boxes" (p717) with some perceiving that less attention was now paid to non-incentivised conditions.[37] Interestingly, patients also perceived that the latter was a potential risk.[38]

Paralleling these organisational changes, GPs and nurses claimed that consultation with patients had become increasingly focused on QOF incentivised care which inevitably had a biomedical focus (tunnel vision).[27, 32, 37, 39] Studies included observation of QOF review consultations alongside interviews with clinicians and patients concluded that such consultations were dominated by a biomedical focus with little room for the patient voice, partly driven by the use of computerised templates during consultations.[40, 41] A consistent finding was concern that recall to multiple disease-focused clinics risked fragmenting care and reducing continuity in terms of patients ability to see 'their' GP, a concern that was shared by patients.[38] It is important to recognise that such concerns predate QOF reflecting that more systematic, more specialised disease management was already happening (and likely contributed to the pre-QOF improvements in quality),[42, 43] However, using survey data from 2003, 2005 and 2007, Campbell et al found that patient experience did not change across most domains (including patient perceptions of communication with GPs and nurses), except that there were small improvements in experience of access for people with chronic illness but not the general population, and for both groups large decreases in the experience of continuity of care (6.7 percentage point reduction in the percentage of people saying they could see their usual doctor, and a 4.5 percentage point reduction in the percentage of people satisfied with this).[44]

Overall, multiple studies found concerns about increased routinisation of care and the implications for professional autonomy and patient care, with some observational evidence that this translated into a biomedical focus in QOF reviews. However, this was in the context of broad approval for most (but not all) QOF indicators because the care incentivised was perceived as important, and a degree of pride in delivering QOF care to high standards. Most professionals described making QOF 'fit in' with their existing way of doing things,[31, 45] even though most researchers perceived the organisational changes as more radical. Two consistent concerns related to worsening continuity of care (for which there was quantitative evidence) and with the crowding out of non-incentivised care.

Crowding out

Many of the participants in the qualitative studies expressed concern that the tunnel vision of a focus on QOF measured care was sometimes to the detriment of unmeasured care. This is an inevitable problem in any quality system based around measurement, since what can be measured is typically only a small proportion of total care. In the QOF context, incentivised conditions only ever represented about 15-20% of workload, and QOF indicators only measured some aspects of care for incentivised conditions. Crowding out of care for other conditions was identified as a risk by both clinicians[46] and patients,[38] and there is some quantitative evidence that it occurred. Steel et al examined changes in quality of care before (2003) and after (2005) QOF. There were large improvements in overall quality measured by directly incentivised indicators, smaller improvements for other indicators for incentivised conditions, and no change in quality measured by indicators for non-incentivised conditions.[47] Analysing data over a longer period, others found that quality of care for almost all indicators was improving pre-QOF, but that QOF was associated with an increased rate of improvement for incentivised indicators and a slowing in the rate of improvement for non-incentivised.[6, 7] This is consistent with practices diverting resources towards the delivery of incentivised care at the partial expense of non-incentivised (although in a resource constrained context, any quality improvement programme is likely to have similar effects).

Gaming

All target driven and financially incentivised measurement will be subject to gaming, which lies in the space between legitimate maximising of performance (for example, by meticulous recording of care being given) and fraud (for example, by false recording). When measurement is high-stakes, then better recording of care that is already being given is to be expected (which will boost the apparent impact of QOF implementation, as appears likely to have happened with the smoking indicators[9-11]), but this will shade over into gaming. Some researchers detected what they interpreted to be gaming in relation to the recording of an unexpectedly high proportion of blood pressure measurements just below targets triggering payment and a deficit just above,[48] although the absolute differences were small and others found no such effect.[49] More convincingly and consistent with qualitative data suggesting that GPs were more willing to exception report when they didn't perceive incentivised care to have much value,[33, 50] the introduction of highly contested indicators incentivising the assessment of depression severity at diagnosis led to a small shift away from the use of QOF-qualifying codes for depression which triggered the requirement to assess severity to symptomatic codes for low mood which did not.[51]

A particular focus of concern about gaming in QOF related to exception reporting. Practices were allowed to remove patients from measurement for a range of reasons including patients not responding to three invitations to attend review, patients declining to take offered treatment, unsuitability because of frailty or terminal illness, patients already being on maximal treatment, and allergy or other strong contraindication. Exception reporting was variously perceived as essential to reduce perverse incentives to either give people futile treatment or deregister the non-compliant, or as an invitation to cheat. Overall though, exception reporting rates were fairly low on average.[52] A median of 5.3% of patients were exception reported, which increased practice income by a median of 2.8%.[52]

Exception reporting rates were lower for simple process indicators like recording of smoking status, and higher rates for treatments where suitability and patient preference will be more important.[52] Those exception reported were on average older and had more comorbidities consistent with acceptable exclusion.[53, 54] More troublingly, rates of exception reporting varied widely between practices with high rates in a few[52] and ethnic minorities and those living in more deprived areas were somewhat more likely to be exception reported.[53]

Gaming is an inevitable part of any measurement system, although measurement design can and should seek to minimise it. Overall, although a few practices had suspiciously high exception reporting rates (which was planned for in QOF through annual inspection and if necessary withholding of payment), exception reporting did not appear to be commonly used to game the payment system. There is some evidence of gaming of specific measures, the most convincing of which is under-recording of depression using QOF-qualifying codes although even here the impact is small in absolute terms.[51]

Implementation failures in QOF

QOF was so large and complicated that implementation failures were inevitable, but several of these have implications for any quality improvement programme and so are briefly reviewed here. Box 2 summarises some of the implementation failures.

Winning hearts and minds

Despite some initial misgivings and concerns about workload and opportunity costs, QOF retained broad professional support for most of its elements throughout its life. This reflected that the core indicators with the most points were for care which clinicians believed was important not least because QOF came after a decade or more of guideline development and dissemination, and quality improvement activity including national service frameworks and clinical audit activity. Some later indicators lacked this belief, notably the depression indicators introduced in 2006 and withdrawn in 2013, which incentivised the screening of patients with coronary heart disease and diabetes for depression, and depression severity assessment at diagnosis using a structured questionnaire. There was good evidence that many GPs and nurses never accepted that the specific care for depression being incentivised mattered[55, 56] (partly reflecting a lack of evidence, but partly reflecting that there was no systematic effort to persuade or educate), and compliance with indicators was often perfunctory or actively gamed (one of the few clear examples of gaming, consistent with the wider qualitative evidence that GPs considering gaming less problematic if they didn't value the care incentivised[33]).[51, 55, 57]

The key lesson is that measurement, target setting and payment may not work and/or are more likely to have perverse consequences if they are not aligned to internal motivation, and so attention to persuading clinicians about importance and value to win hearts and minds is essential. More broadly, there are no magic bullets in quality improvement, and success will usually require deploying multiple strategies which might include education, persuasion, feedback, public reporting, facilitation of change, and payment.[58]

Technical problems with indicators

Measuring quality is difficult and easy to get wrong. Several indicators have required revision or been withdrawn because of technical problems including there simply not being enough patients per practice to reliably measure quality (an issue affecting many indicators in small practices, whose measured quality and therefore payment varied through the play of chance[15, 59]). Others have fallen during piloting because they turn out not to be feasible either to measure or to implement, including for example proposed palliative care indicators relating to recording preferred place of death.[46, 60] The issue with the latter wasn't that the care incentivised wasn't perceived to be important, but that paying to routinely deliver such a complex type of care carried serious risks of perverse consequences as well as being subject to small numbers problems.

Some were implemented with the best of intentions, but had perverse effects, notably the change in April 2006 to the Read Codes used to define the diabetes register were changed, with the intention of ensuring that the type of diabetes was recorded in the electronic medical record. However, the actual outcome was to remove people with unspecified diabetes Read Codes from the QOF register and therefore from at least some recall systems, with evidence that care was systematically worse for patients removed from the register.[3] Finally, indicators created without adequate scrutiny and piloting were frequently problematic including the obesity register whose implementation bypassed the National Institute for Health and Care Effectiveness process of indicator design and piloting. Its creation was in response to legitimate government concerns that obesity was a major problem that required action, but in the face of a lack of evidence of effective primary care interventions. What was implemented was an "obesity register" which like all other registers simply paid practices the full incentive once they recorded one person as having the condition (in this case a body mass index of 30 or more). There was therefore no actual incentivisation of a change in practice (since every practice would have at least one such person with a high body mass index recorded). If the intention was to encourage regular measurement of weight and height, then this would have been straightforward to incentivise in the same way that practices were incentivised to regularly measure blood pressure or record smoking status in adults. An obesity register would then have been created by more systematic measurement, rather than the waste of money which actually happened.

The key lesson is that the design of quality measures is complicated, and minimising perverse effects is best achieved by close cooperation between policymakers, technical experts and clinicians to ensure that indicators are fit for the purpose to hand.[61]

Payment system problems

The basis on which negotiators allocated points (and therefore the level of payment) to particular indicators was never that explicit, although it broadly appeared to reflect workload. One criticism of it was that the it was that the allocation was poorly linked to health gain,[62, 63] and even if the intention was to match workload, the same clinical action was rewarded differently depending on the condition a patient had (with for example, more than two-fold variation in the payment for controlling one patient's blood pressure.[64] Most problematically, the original payment system would have paid practices in relation to their capitation (global sum) weighting which was widely criticised in the year before implementation for failing to reflect the prevalence of disease. An 'adjusted disease prevalence factor' was therefore fairly hurriedly introduced but amended to protect the income of practices with low prevalence of disease. Unfortunately, there was an error in its implementation which was not corrected for six years, and meantime effectively paid larger practices more for delivering the same level of quality as smaller practices, and systematically diverted resource from more deprived to more affluent practices.[65]

The key lesson is that payment systems should wherever possible be modelled in real data before implementation, and any errors in them rapidly corrected.[65]

Inadequate attention to evaluation

Implementing the impact of QOF was always going to be problematic because it was implemented across the whole of the UK simultaneously, but opportunities to address important later questions were missed,

Box 2: Examples of implementation failures in QOF

Winning hearts and minds

The initial set of clinical indicators retained broad professional support throughout the lifetime of QOF, at least partly because clinicians had been persuaded of their value after a decade or more of guideline development and national service frameworks. This was not true of all later indicators, notably the depression indicators introduced in 2006 and withdrawn in 2013, where GPs were unpersuaded of their clinical value, and compliance was variable and sometimes perfunctory.

The key lesson is that measurement, target setting and payment may not work or may be more likely to have perverse consequences if they are not aligned to internal motivation, so attention to persuading clinicians about important and value is essential.

Technical problems with indicators

Many indicators had small numbers of patients being measured in each practice, which led to payment to small practices varying considerably due to chance variation.

The Read Codes used to define the diabetes register were changed in April 2006 to only include codes specifying the type of diabetes in order to improve register quality. The short-term effect was to exclude patients with unspecified diabetes Read Codes from the QOF register meaning that some of them were lost to routine follow-up and received worse care.

The introduction of the 'obesity register' bypassed piloting. As implemented, it paid practices for having a 'register' of people with body mass index (BMI) >30 kg/m^2, with payment triggered the moment the first such person had their BMI recorded.

The key lessons are that even with the best of intentions, indicators can have perverse consequences (most of which can be picked up in piloting) and design needs close cooperation between policymakers, technical experts and clinicians to ensure that there is clarity about what the indicator is intended to achieve (never specified for the obesity register for example) and to maximise indicator fitness for that purpose.

Payment system problems

The initial payment system was rapidly modified in response to initial criticism that payment would not match workload, but the modification included changes to protect practices with low workload and was implemented incorrectly. The outcome was to pay larger practices more than smaller for delivering the same quality of care, and to systematically divert resource from more deprived to more affluent practices. The key lesson is that payment systems should be modelled in real data before implementation.

Inadequate attention to evaluation

There were many missed opportunities for evaluation in QOF, including the failure to collect data about quality of care when indicators were withdrawn, which would have been useful to inform decisions about QOF abolition (in Scotland) and scaling down (in the rest of the UK). *The key lesson* is that decisions about data collection should be informed by consideration of policy-relevant knowledge gaps.

including the impact of exception reporting (where relevant data was not extracted in the first year) and evaluating the effect of withdrawing indicators (which is obviously of great interest currently). Routine reporting of QOF data for withdrawn indicators would have significantly improved our understanding of indicator withdrawal. A single study using a research GP dataset evaluated the effect of indicator withdrawal and concluded that quality did not significantly decline apart from a small reduction in influenza immunisation for people with asthma (withdrawn because of limited evidence of benefit).[66] However, practices contributing to research datasets are atypical in several ways, and the indicators evaluated mostly remained incentivised in some way within QOF. Certainly, one of the few other studies of incentive withdrawal found that quality declined, in some cases below the level when incentives had originally been introduced.[46]

The key lesson is that decisions about data collection should be informed by careful consideration of policy-relevant knowledge gaps.

Conclusions

The Quality and Outcomes Framework was a heroic endeavour: an unprecedentedly large pay-for-performance programme in terms of being nationwide, using almost 150 indicators and linking up to 25% of practice income to performance. In terms of the core bargain between government and GPs, in the short-term it delivered higher performance to the government (in relation to incentivised quality of care) and higher pay to the GPs, with an improvement in the shared problem of low morale. Over time, quality has remained high (but hasn't clearly further improved, partly because of ceiling effects), but practice income and pay has declined, and GPs' satisfaction with their work is back to pre-2004 levels, consistent with the original bargain breaking down. QOF had a number of unintended consequences and technical failures of implementation, which provide several key lessons for any future large-scale improvement programme.

Appendix 1: QOF design in brief

Most payment systems are complicated and QOF was no exception. The 'pay' bit of QOF was points, where each point was worth a certain amount of money. How much it was worth varied between practices depending on their prevalence of the diseases for which care was incentivised, but all practices could earn up 1000 points each year (with an additional 50 point 'access bonus' initially). The maximum total payment represented 20-25% of practice income, and was new money so represented a substantial increase. Points could be earned in a number of organisational domains (eg medicines management, patient experience) and a number of clinical domains (eg coronary heart disease, diabetes, severe mental illness, cancer). The 'performance' side was represented by 147 indicators initially with the number of indicators in each domain varying from one (eg obesity which was only ever a payment for having a register) to 18 (for diabetes). Some indicators were all or nothing in that practices either earned the points or not, whereas others were paid on a sliding scale with nothing earned until performance crossed a minimum threshold, then increasing payment until a maximum performance threshold which was 90% for almost all process indicators but lower for most intermediate outcome indicators. For example, in the first year QOF (2004/5):

- MED09 A medication review is recorded in the notes in the preceding 15 months for all patients being prescribed repeat medicines (excluding OTC and topical medications): Standard 80 per cent. This is an all or none indicator, with practices earning eight points once they cross the 80% threshold.
- DM 6. The percentage of patients with diabetes in whom the last HbA1C is 7.4 or less (or equivalent test / reference range depending on local laboratory) in last 15 months. This is a sliding scale indicator, where practices earned nothing until at least 25% of patients achieved this level, and would earn all 16 points once 50% did so.

Over time, thresholds increased, with the minimum threshold rising to 40% for all indicators, and maximum thresholds being progressively raised. Indicators were also retired for a variety of reasons including to make room for new indicators which were initially produced by a consortium of professionals and latterly by the National Institute for Health and Care Excellence.[1,67]

References

1. Roland M, Campbell S. Successes and Failures of Pay for Performance in the United Kingdom. *New England Journal of Medicine* 2014; **370**(20): 1944-9.
2. British Medical Association. Investing in General Practice: the New General Medical Services Con tract. London: Brit ish Medical Association, 2003.
3. Calvert M, Shankar A, McManus RJ, Lester H, Freemantle N. Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study. [Erratum appears in BMJ. 2009;339:b2768]. *Bmj* 2009; **338**: b1870.
4. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: A longitudinal observational study. *BMJ Quality and Safety* 2013; **22**(1): 53-64.
5. Serumaga B, Ross-Degnan D, Avery AJ, et al. Effect of pay for performance on the management and outcomes of hy pertension in the United Kingdom: inter rupted time series study. *Bmj* 2011; **342**: d108.
6. Doran T, Kontopantelis E, Valderas JM, et al. Effect of finan cial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework.[Erratum appears in BMJ. 2013;347:f5939]. *Bmj* 2011; **342**: d3590.
7. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Ro land M. Effects of Pay for Performance on the Quality of Primary Care in England. *New England Journal of Medicine* 2009; **361**: 368-78.
8. Arrowsmith ME, Majeed A, Lee JT, Saxena S. Impact of pay for performance on prescribing of long-acting reversible contraception in primary care: an interrupted time series study. *PloS One* 2014; **9**(4).
9. Taggar JS, Coleman T, Lewis S, Szatkowski L. The impact of the Quality and Outcomes Framework (QOF) on the re cording of smoking targets in primary care medical rec ords: cross-sectional analyses from The Health Improve ment Network (THIN) database. *BMC Public Health* 2012; **12**(pp 329).

References (continued)

10. Szatkowski L, Aveyard P. Provision of smoking cessation support in UK primary care: impact of the 2012 QOF revi sion. *British Journal of General Practice* 2016; **66**(642): e10-5.
11. Coleman T. Do financial incentives for delivering health promotion counselling work? Analysis of smoking cessation activities stimulated by the quality and outcomes framework. *BMC Public Health* 2010; **10**(pp 167).
12. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for perfor mance scheme on emergency hospital admissions for am bulatory care sensitive conditions: controlled longitudinal study. *Bmj* 2014; **349**(pp g6423).
13. Fleetcroft R, Parekh-Bhurke S, Howe A, Cookson R, Swift L, Steel N. The UK pay-for-performance programme in primary care: estimation of population mortality reduction. *British Journal of General Practice* 2010; **DOI: 10.3399bjgp10X515359**: e345.
14. Ryan AM, Krinsky S, Kontopantelis E, Doran T. Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study. *The Lancet* 2016; **388**(10041): 268-74.
15. Doran T, Campbell S, Fullwood C, Kontopantelis E, Roland M. Performance of small general practices under the UK's Quality and Outcomes Framework. *British Journal of Gen eral Practice* 2010; **60**(578): e335-e44.
16. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *The Lancet* 2008; **372**(9640): 728-36.
17. Norbury M, Fawkes N, Guthrie B. Impact of the GP contract on inequalities associated with influenza immunisation: A retrospective population-database analysis. *British Journal of General Practice* 2011; **61**(588): e379-e85.
18. Millett C, Saxena S, Netuveli G, Majeed A. Impact of Pay for Performance on Ethnic Disparities in Intermediate Out comes for Diabetes: A Longitudinal Study. *Diabetes Care* 2009; **32**: 404-9.
19. Lee JT, Netuveli G, Majeed A, Millett C. The effects of pay for performance on disparities in stroke, hyper tension, and coronary heart disease management: interrupted time series study. *PLoS ONE [Electronic Resource]* 2011; **6**(12): e27236.
20. Royal College of General Practitioners. A blueprint for Scottish general practice: A strategy for a safe, secure and strong general practice in Scotland. Edinburgh, Scotland: Royal College of General Practitioners Scotland, 2015.
21. Lafond S. Current NHS spending in England. London: The Health Foundation, 2015.
22. Gemmell I, Campbell S, Hann M, Sibbald B. Assessing workload in general practice in England before and after the introduction of the pay-for-performance contract. *Journal of Advanced Nursing* 2009; **65**(3): 509-15.
23. Health and Social Care Information Centre. GP Earnings and Expenses Time Series 2002/03 to 2012/13 http:// www.hscic.gov.uk/searchcatalogue?productid=13317. 2015.
24. Whalley D, Gravelle H, Sibbald B. Effect of the new contract on GPs' working lives and perceptions of quality of care: a longitudinal survey. *British Journal of General Practice* 2008; **58**: 8-14.
25. Gibson J, Checkland K, Coleman A, et al. Eighth National GP Worklife Survey. Manchester: Policy Research Unit in Commissioning and the Healthcare System, University of Manchester, 2016.
26. Campbell SM, McDonald R, Lester H. The experience of pay for performance in english family practice: A qualitative study. *Annals of Family Medicine* 2008; **6**(3): 228-34.
27. Checkland K, Harrison S, McDonald R, Grant S, Campbell S, Guthrie B. Biomedicine, holism and general medical practice: Responses to the 2004 General Practition er contract. *Sociology of Health and Illness* 2008; **30**(5): 788 -803.
28. Mangin D, Toop L. The quality and outcomes frame work: What have you done to yourselves? *British Journal of General Practice* 2007; **57**(539): 435-7.
29. Toop L. The QOF, NICE, and depression: a clumsy mechanism that undermines clinical judgment. *British Journal of General Practice* 2011; **61**(588): 432-3.
30. Alyahya M. Changing organizational structure and organizational memory in primary care practices: A qualitative interview study. *Health Services Management Research* 2012; **25**(1): 35-40.
31. Checkland K, Harrison S. The impact of the Quality and Outcomes Framework on practice organisation and ser vice delivery: summary of evidence from two qualita tive studies. *Quality in Primary Care* 2010; **18**(2): 139-46
32. McDonald R, Harrison S, Checkland K, Campbell SM, Ro land M. Impact of financial incentives on clinical autonomy and internal motivation in primary care: Ethnographic study. *British Medical Journal* 2007; **334**(7608): 1357-9.
33. Maisey S, Steel N, Marsh R, Gillam S, Fleetcroft R, Howe A. Effects of payment for performance in primary care: Qualita tive interview study. *Journal of Health Services Re search and Policy* 2008; **13**(3): 133-9.
34. Lester H, Matharu T, Mohammed MA, Lester D, Foskett-Tharby R. Implementation of pay for performance in primary care: A qualitative study 8 years after introduction. *British Journal of General Practice* 2013; **63**(611): e408-e15.
35. Cheraghi-Sohi S, Calnan M. Discretion or discretions? De lineating professional discretion: the case of English medical practice. *Social Science & Medicine* 2013; **96**(pp 52-59).
36. Grant S, Huby G, Watkins F, et al. The impact of pay-for-performance on professional boundaries in UK general practice: an ethnographic study. *Sociology of Health and Illness* 2009; **31**(2): 229–45.
37. McGregor W, Jabareen H, O'Donnell CA, Mercer SW, Watt GCM. Impact of the 2004 GMS contract on practice nurs es: A qualitative study. *British Journal of General Practice* 2008; **58**(555): 711-9.
38. Hannon KL, Lester HE, Campbell SM. Patients' views of pay for performance in primary care: A qualitative study. *British Journal of General Practice* 2012; **62**(598): e322-e8.
39. McDonald R, Checkland K, Harrison S. The new GP con tract in English primary health care: an ethnographic study. *International Journal of Public Sector Management* 2009; **22** (1): 21-34.
40. Chew-Graham CA, Hunter C, Langer S, et al. How QOF is shaping primary care review consultations: a longitudinal qualitative study. *BMC Family Practice* 2013; **14**(pp 103).
41. Blakeman T, Chew-Graham C, Reeves D, Rogers A, Bower P. The Quality and Outcomes Framework and self-management dialogue in primary care consultations: A qual itative study. *British Journal of General Practice* 2011; **61** (591): e666-e73.
42. Guthrie B, Wyke S. Access and continuity in UK general practice: a qualitative study of general practitioners' and patients' perceptions of when and how they matter. *BMC Family Practice* 2006; **7** 11.
43. Charles-Jones H, Latimer J, May C. Transforming general practice: the redistribution of medical work in primary care. *Sociology of Health & Illness* 2003; **25**(1): 71-92.
44. Campbell SM, Kontopantelis E, Reeves D, et al. Changes in patient experiences of primary care during health service reforms in England between 2003 and 2007. *Annals of Fam ily Medicine* 2010; **8**: 499-506.
45. Huby G, Guthrie B, Grant S, et al. Whither British general practice after the 2004 GMS contract?: Stories and realities of change in four UK general practices. *Journal of Health Organisation and Management* 2008; **22**: 63-78.
46. Lester HE, Hannon KL, Campbell SM. Identifying unintend ed con sequences of quality indicators: A qualitative study. *BMJ Quality and Safety* 2011; **20**(12): 1057-61.
47. Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: an observational study. *British Journal of General Practice* 2007; **57**: 449-54.
48. Carey IM, Nightingale CM, DeWilde S, Harris T, Whincup PH, Cook DG. Blood pressure recording bias during a peri od when the Quality and Outcomes Framework was intro duced. *Journal of Human Hypertension* 2009; **23**(11): 764-70.
49. Karunaratne K, Stevens P, Irving J, et al. The impact of pay for performance on the control of blood pressure in people with chronic kidney disease stage 3-5. *Nephrology Dialysis Transplantation* 2013; **28**(8): 2107-16.
50. Campbell S, Hannon K, Lester H. Exception reporting in the quality and outcomes framework: Views of practice staff - A qualittive study. *British Journal of General Practice* 2011; **61** (585): e183-e9.

References (continued)

51. Kendrick T, Stuart B, Newell C, Geraghty AW, Moore M. Changes in rates of recorded depression in English primary care 2003-2013: Time trend analyses of effects of the eco nomic recession, and the GP contract quality outcomes framework (QOF). *Journal of Affective Disorders* 2015; **180** (pp 68-78).
52. Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *The New England Journal of Medicine* 2008; **359**: 274-84.
53. Dalton ARH, Alshamsan R, Majeed A, Millett C. Exclusion of patients from quality measurement of diabetes care in the UK pay-for-performance programme. *Diabetic Medicine* 2011; **28**(5): 525-31.
54. Simpson C, Hannaford P, McGovern M, et al. Are different groups of patients with stroke more likely to be excluded from the new UK general medical services contract? A cross -sectional retrospective analysis of a large primary care pop ulation. *BMC Family Practice* 2007; **8**: 56.
55. Leydon GM, Dowrick CF, McBride AS, et al. Questionnaire severity measures for depression: A threat to the doctor-patient relationship? *British Journal of General Practice* 2011; **61**(583): 117-23.
56. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and manage ment of depression: A qualitative study. *British Journal of General Practice* 2011; **61**(586): e279-e89.
57. Dowrick C, Leydon GM, McBride A, et al. Patients' and doc tors' views on depression severity questionnaires incentive ised in UK quality and outcomes framework: Qualitative study. *BMJ: British Medical Journal* 2009; **338**(7697): 1-9.
58. Roland M, Dudley RA. How Financial and Reputational In centives Can Be Used to Improve Medical Care. *Health Serv Res* 2015; **50**(Supplement S2): 2090-115.
59. Roland M, Elliott MN, Lyratzopoulos G, et al. Reliability of patient responses in pay for performance schemes: analysis of national General Practitioner Patient Survey data in Eng land. *Bmj* 2009; **339**: h3851.
60. Hannon KL, Lester HE, Campbell SM. Recording patient preferences for end-of-life care as an incentivized quality indicator: What do general practice staff think? *Palliative Medicine* 2012; **26**(4): 336-41.
61. Lester H, Campbell S. Developing Quality and Outcomes Framework (QOF) indicators and the concept of 'QOFability'. *Quality in Primary Care* 2010; **18**(2): 103-9.
62. Fleetcroft R, Cookson R. Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Services Research and Poli cy* 2006; **11**(1): 27-31.
63. Fleetcroft R, Steel N, Cookson R, Walker S, Howe A. Incen tive payments are not related to expected health gain in the pay for performance scheme for UK primary care: cross-sectional analysis. *BMC Health Serv Res* 2012; **12**(pp 94).
64. Ruscitto A, Mercer SW, Morales D, Guthrie B. Accounting for multimorbidity in pay for performance: a modelling study using UK Quality and Outcomes Framework data. *British Journal of General Practice* 2016.
65. Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general prac tice contract. *British Journal of General Practice* 2006; **56** (532): 836-41.
66. Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Val deras JM, Doran T. Withdrawing performance indicators: retrospective analysis of general practice performance un der UK Quality and Outcomes Framework.[Erratum appears in BMJ. 2014;349:g7730]. *Bmj* 2014; **348**: g330.
67. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. *N Engl J Med* 2004; **351**: 1448 - 54.