The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review

Carwyn Langdown1, Stephen Peckham2,3

1Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
2Centre for Health Services Studies, University of Kent, George Allen Wing, Canterbury, Kent CT2 7NF, UK
3Policy Research Unit in Commissioning and the Healthcare System, University of Manchester and CHSS, UK
Address correspondence to Carwyn Langdown, E-mail: cjlangdown@hotmail.com

ABSTRACT

Background The quality and outcomes framework (QOF) is one of the world’s largest pay-for-performance schemes, rewarding general practitioners for the quality of care they provide. This review examines the evidence on the efficacy of the scheme for improving health outcomes, its impact on non-incentivized activities and the robustness of the clinical targets adopted in the scheme.

Methods The review was conducted using six electronic databases, six sources of grey literature and bibliography searches from relevant publications. Studies were identified using a comprehensive search strategy based on MeSH terms and keyword searches. A total of 21 543 references were identified of which 32 met the eligibility criteria with 11 studies selected for the review.

Results Findings provide strong evidence that the QOF initially improved health outcomes for a limited number of conditions but subsequently fell to the pre-existing trend. There was limited impact on non-incentivized activities with adverse effects for some sub-population groups.

Conclusion The QOF has limited impact on improving health outcomes due to its focus on process-based indicators and the indicators’ ceiling thresholds. Further research is required to strengthen the quality of evidence available on the QOF’s impact on population health to ensure that the incentive scheme is both clinically and cost-effective.

Keywords indicators, primary care, QOF

Introduction

The quality and outcomes framework (QOF) was introduced in the UK in 2004 as part of the new general practitioner contract. The aim of the QOF was to provide ‘a major focus on quality and outcomes’, incentivizing practices to achieve higher standards of quality care.1

As one of the world’s largest pay-for-performance schemes of its kind, at a cost of approximately £1 billion a year,2 the QOF provides GPs an opportunity to increase practice income through a points-based system. These are awarded when a range of indicators are met, subject to attaining a minimum and maximum level of thresholds and targets. Indicators are based on four different domains: clinical, organizational, patient experience and additional services, with a maximum of 1000 points currently available across the domains, which determine the amount each practice is paid, with each point valued to be £133.76.3

The QOF has five different types of clinical indicators. Both ‘health outcome’ and ‘intermediate outcome’ measures are the main indicators that measure any real impact on health. The ‘health outcome’ indicator (of which there is only one) measures the number of epilepsy patients, which have been seizure free in the past 12 months. The ‘intermediate outcome’ indicator measures outcomes such as the achievement of specific blood pressure and cholesterol targets, i.e. indicators that measure changes in health status that affect health outcomes. The remaining three clinical indicator types are more process based, focused on...
measuring activities such as the number of blood pressure readings which may be directly or indirectly associated with health outcomes. These include ‘process measures directly linked to health outcomes’; ‘process measures’ and ‘register’.

In its first year, 222 (2.6%) practices in England achieved the maximum scores available of 1050 points (maximum available has now reduced to 1000 points), with almost half the practices achieving a score between 1000 and 1050.4 So does a high QOF score represent high quality care and better outcomes, or does it reflect better housekeeping with practices becoming better at recording patient activity?

Gap in research

There is a substantial literature on QOF; however, few studies examine the efficacy of the scheme for improving health outcomes. While some studies focus on specific conditions such as diabetes,5–9 few have taken a broader approach, which considers the impact on a range of different conditions and the unintended consequences on non-incentivized aspects of care, which this review seeks to address. Similarly, few have explored how robust the indicators clinical target thresholds are when compared with clinical best practice such as National Institute for Health and Clinical Excellence (NICE) guidelines, which is also considered in this review.

Methods

Six electronic databases were searched for publications, selected based on their subject coverage and applicability to this review—MEDLINE, EMBASE, Cochrane, Econlit, Web of Science and Health Management Information Consortium. Grey literature was also included to identify material from organizations including the NICE; The NHS Information Centre, Department of Health, The King’s Fund; NHS Employers; World Health Organization.

Search terms were drawn from three broad concepts—primary care; financial incentives and health outcomes. Each concept was used to identify associated keywords, related synonyms and MeSH terms. Boolean operators ‘OR’ and ‘AND’ were used to combine terms and determine how the results would be integrated, such as ‘QOF’ OR ‘financial incentives’ (see Supplementary data, Information 2: Search Terms Strategy & Results).

Studies were restricted to those published from January 2004 (the year the QOF commenced) to June 2012 and limited to English language publications focusing entirely on the QOF’s clinical indicators. Qualitative studies were excluded as quantitative evidence was deemed more appropriate in demonstrating an improvement or decline in the achievement of health outcomes. Supplementary data, Information 1 illustrates the inclusion and exclusion criteria adopted in this review. Given the broad nature of the term ‘health outcomes’, a range of outcome measures were developed for the review based on the QOF’s outcome and intermediate outcome indicators (see Supplementary data, Information 3: Outcome Measures).

A total of 21 354 references were identified as illustrated in Fig. 1 search process and corresponding results. Following a review of the abstracts 44 papers were selected with an additional 5 references identified from the grey literature search. Full text copies of these papers were retrieved and, together with two further references identified from a bibliography search, assessed to identify whether they met the inclusion criteria for the review.

Ultimately 32 papers met the inclusion criteria and these were assessed for quality based on the Critical Appraisal Skills Programme and Cochrane’s EPOC checklist assessment tools.10,11 Twenty-one studies were discounted during this exercise, failing to meet the quality standards due to their methodological quality and strength of findings such as the absence of pre-QOF data or short follow-up period (see Supplemental data, Information 4: Quality Assessment).

Results

Eleven studies met the quality criteria for inclusion in the review. The following presents the results of the studies analysed by the reviews’ core objectives: impact on health outcomes; non-incentivized activities and clinical target thresholds.

Impact of QOF on health outcomes

Ten of the 11 studies selected looked at the impact of QOF on various health outcomes (n = 10)5–9,12–16 of which, 7 focused on one specific condition. All seven studies, which included diabetes within the research, reported a significant increase in the achievement of intermediate health outcome targets (blood pressure ≤145/85 mmHg; cholesterol <5 mmol/l; HbA1c ≤ 7.4%). Hippisley-Cox et al.16 reported a 56% relative increase in the percentage of diabetic patients with the blood pressure ≤145/85 mmHg in the previous 15 months to 74% in 2006. Achievement of cholesterol targets reported an even higher relative increase of 132%, with 74% of patients with cholesterol <5 mmol/l in 2006. However, the authors failed to quantify the increase in targets achieved in relation to the underlying trend, which existed prior to the introduction of QOF.

The results presented by Campbell et al.16 did consider the pre-QOF trend, which existed and demonstrated that the clinical-quality scores for diabetes were improving before the introduction of QOF at an average rate of 1.8% per annum. This increased significantly in 2005 after the introduction of QOF, above the underlying trend, although fell to the pre-existing trend level beyond 2005.
Campbell et al.\textsuperscript{16} also explored the impact of QOF on other conditions including asthma and CHD. It demonstrated that the rate of improvement in the mean clinical-quality scores for asthma increased between 2003 and 2005, above the increasing trend which existed before the introduction of QOF, although this was not maintained after 2005. For CHD,
a slight increase occurred in 2005 following the introduction of QOF, but was not significantly higher than the underlying trend present before the QOF was implemented.

Few studies were identified for epilepsy (the only ‘health outcome’ indicator) that were relevant to the review’s objectives. However, these were excluded due to the inclusion/exclusion criteria or their methodological quality.

**Impact on non-incentivized clinical activities**

Two studies considered the impact of the QOF on non-incentivized clinical activities and associated health outcomes ($n = 2$). Doran et al. examined whether the introduction of QOF led to activities not included in the pay-for-performance scheme to be neglected. During the QOF’s first year, achievement in incentivized indicators increased above the existing trend but reached a plateau by 2006/07. No statistically significant changes occurred in the achievement of non-incentivized indicators above the pre-existing trend during the QOF’s first year. However by 2006/07, these declined to a level that was lower than predicted and below the underlying trend for non-incentivized indicators.

Campbell et al. considered the impact on non-incentivized clinical activities. The QOF’s impact on asthma for non-incentivized activities during its first year was not significantly different from activities that were incentivized. However, trends for non-incentivized care declined after 2005 where an increase occurred for incentivized care. After 2005, the mean quality score for un-incentivized activities declined, but increased for indicators that were incentivized (Fig. 2).

**QOF’s clinical target thresholds**

One study examined the robustness of the clinical target thresholds set within the QOF’s clinical indicators compared with best practice standards ($n = 1$). Strong et al. assessed the quality of the spirometry measures related to the QOF COPD indicators against the British Thoracic Society standards. It demonstrated that the proportion of patients whose spirometry met BTS standards amounted to 31% of cases. Also, 12% of patients on the COPD register had FeV1 (%) predicted that did not support the diagnosis of COPD. It concluded that there was no association between the quality measured by BTS standards and the achievement of QOF COPD indicators (COPD9 and 10), citing that the QOF assesses the quantity rather than the quality of spirometry.

The clinical threshold targets for a selection of QOF outcome and intermediate outcome indicators, were compared against the evidence-based clinical guidelines cited within the QOF guidance, which provided the rationale for what the target would be set. As illustrated in Table 1 (indicator differences between QOF thresholds and evidence-based guidelines), 4 of the 16 indicators were identified to have different clinical threshold levels to that recommended within the NICE or Scottish Intercollegiate Guidelines Network (SIGN) clinical guidelines drawn from the QOF guidance. This highlights that the optimal blood pressure targets recommended are stricter in the guidelines than the actual thresholds targets set within the QOF.

**Discussion**

**Main findings of this study**

**Impact of QOF on health outcomes**

The evidence from the results suggests that the QOF has led to an improvement in health outcomes for some conditions including Diabetes, although the results are mixed for others such as CHD. Also, despite a surge of improvement during its
introductory period for some conditions, levels of achievement reached a plateau in later years which may be due to a ceiling effect caused by the maximum threshold levels set for each indicator as practices were not incentivized to improve health outcomes beyond the various clinical target threshold levels (e.g. 85% of practice population). This specifies the minimum and maximum percentages of patients that are required to achieve the target before the minimum and maximum scores are rewarded for that indicator. If incentives are designed to influence behaviour, shouldn’t the achievement thresholds be set at an optimal level? The principles for the review of thresholds were discussed in the QOF advisory committee (2009) where it considered the option of setting the maximum threshold at 100%. However committee members raised the issue that this may create a struggle for practices in deprived areas, leading to a subsequent adverse impact on health inequalities.

The studies highlight that the QOF is currently limited in what it measures in terms of health outcomes. Only one indicator is solely focused on the achievement of an actual health outcome (i.e. the number of Epilepsy patients which have been seizure free in the last 15 months), whereas the remaining intermediate outcomes relate to targets which are an indirect measure of one’s health, e.g. cholesterol<5 mmol/l. The QOF points available are also weighted towards particular conditions such as Diabetes (88 points), and CHD- secondary prevention (69 points); compared to scores available for COPD (30 points) and Depression (31 points).

The Darzi report ‘High Quality Care For All: NHS Next stage review’ (2008), acknowledges the QOF’s limitations, recognizing that the QOF’s incentives focus on long-term conditions as opposed to attempting to prevent the conditions from occurring in the first instance. The report advocates reducing the number of process and organizational indicators, refocusing resources on indicators of prevention and clinical effectiveness. The Marmot Review on health inequalities also suggested that the QOF should place greater emphasis on prevention rather than addressing established diseases. It also highlighted the scheme’s potential of achieving maximum QOF points without covering the practice’s total population, which risked disengaging hard to reach groups and potentially, most in need. Marmot recommended revising the QOF to ‘provide 100% coverage of the quality of care for all patients’.

**Impact on non-incentivized activities**

The evidence from the studies suggests that there was no immediate impact on non-incentivized activities following the introduction of the QOF. However, these activities did decline over the longer term in comparison to the temporal trend which existed, with the exception of Diabetes for which there was no significant change from the trend.

As pointed out by Peckham and Wallace, the size of the incentive must be large enough to influence the clinician’s behaviour. This may imply that the relationship between improved health outcomes and incentivized activities under the QOF may be closer related to the number of QOF points available rather than whether a clinical activity is incentivized or not, particularly given that the scheme is voluntary.

**Robustness of the QOF clinical target thresholds**

The literature’s evidence on the clinical target thresholds set for each indicator is limited as reflected by the number of studies identified in this review. The evidence for Spirometry of COPD patients seems to suggest a disparity between the BTS standards and what is regarded under QOF. However, the specific nature of this study prevents its generalizability of the findings to other indicators.

The comparison made between the recommended clinical guideline thresholds (NICE and SIGN) and the QOF

<table>
<thead>
<tr>
<th>Code</th>
<th>Indicator description</th>
<th>Unit of measurement</th>
<th>Recommended target per QOF guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPS</td>
<td>% of patients with hypertension whom the last blood pressure (measured in the previous 9 months) is 150/90 or less</td>
<td>Blood pressure (mmHg)</td>
<td>&lt;140/90 mmHg</td>
</tr>
<tr>
<td>CHD6</td>
<td>% of patients with CHD whom the last blood pressure reading (measured in the previous 15 months) is 150/90 or less</td>
<td>Blood pressure (mmHg)</td>
<td>&lt;140/90 mmHg</td>
</tr>
<tr>
<td>DM30</td>
<td>% of patients with diabetes in whom the last blood pressure reading is 150/90 or less</td>
<td>Blood pressure (mmHg)</td>
<td>&lt;140/90 mmHg</td>
</tr>
<tr>
<td>Stroke6</td>
<td>% of patients with a history of TIA or stroke whom the last blood pressure reading (measured in the previous 15 months) is 150/90 or less</td>
<td>Blood pressure (mmHg)</td>
<td>&lt;140/85 mmHg</td>
</tr>
</tbody>
</table>
indicators, suggests that the blood pressure indicators for the conditions highlighted in Table 1: indicator differences between QOF thresholds and evidence-based guidelines appear to be stricter under the recommended clinical guidelines. The QOF has taken the approach of adopting an ‘audit standard’, for patients of all ages, adhering to the guidelines recommended for those aged over 80, which may have been for pragmatic reasons. However, this leaves a suboptimal clinical target for a large cohort of patients with those conditions and created an opportunity for practices whose demographics for those conditions favour those aged under 80 years. In August 2012, NICE published its QOF recommendations for 2013/14 which proposed lowering these targets for those aged under 80 years in line with clinical guidelines.

What is already know on this topic
While extensive literature exists on the QOF; much of this is focused on a limited number of chronic conditions: diabetes, CHD and hypertension. As the only condition to have an ‘outcome’ indicator, it is surprising that the evidence on the QOF’s impact on epilepsy is limited. However, the read codes for epilepsy did not exist prior to 2003, hence any increases in the achievement of this indicator are likely to be due to an increase in recording than actual outcomes, which may also explain the limited literature available.

What this study adds
The review’s strengths lie in its comprehensive approach, both in the range of data sources used and the breadth of search terms applied within the search strategy. It also considers a range of conditions and health outcomes rather than the specific conditions often found within the research on QOF.

Limitations of review
Given the volume of QOF literature available, it is surprising that only 11 studies were selected. While this may reflect the low quality of studies on the subject, it may also suggest that the review’s methods and selection criteria could have been too specific, excluding studies which may have contributed to the findings. Another limitation of the study is that the review was conducted by one individual, which may also have impacted on the selection of abstracts and the quality assessment conducted.

Conclusion
This review highlights that the availability of high-quality evidence to evaluate the QOF’s impact on health outcomes is limited, mainly due to methodological quality with few providing pre- and post-QOF comparisons with a sufficient follow-up period to evaluate the impact on health outcomes, and the narrow range of conditions included other than diabetes, CHD and hypertension.

The evidence demonstrates that the QOF initially improved blood pressure, cholesterol and Hba1c for diabetes, and the limited impact it has had on other conditions relative to the pre-QOF trend which existed. It also reveals the unintended consequences, which may occur including the non-incentivized clinical activities and the impact on sub-population groups, which may partly be due to the population exclusions permitted.

The review tested the robustness of how the QOF scores are awarded, contributing to the evidence base in this area. It highlights the QOF’s limitations in improving health outcomes, due to the indicators’ ceiling thresholds and the sub-optimal clinical targets when compared with the national clinical guidelines, suggesting that the QOF’s financial incentives may not be sufficiently challenging in some areas.

The evidence demonstrates that the indicators have undoubtedly led to the improvement of disease registers and the recording of clinical activities. The scheme can be used to inform practices of their population’s health needs; however, as Fleetcroft and Cookson note, the incentives operate in a way that rewards practices for ‘high-workload activities’ rather than influencing practices to proactively address health needs and provide preventative services. The evidence also demonstrates that although more practices are achieving higher or maximum QOF points, the ceiling placed on indicator thresholds do not incentivize practices to address the needs of all their population.

Attempts to address the QOF’s weaknesses can be seen at a local level, with health organizations developing their own initiatives to address the scheme’s limitations. Examples include the QOF+ in NHS Hammersmith and Fulham, placing greater emphasis on primary prevention including alcohol screening and interventions, although this comes at an additional cost to the national scheme of £1 million per annum. Current national developments also suggest a shift towards a more outcome-based approach such as the NHS Outcomes Framework and Commissioning Outcomes Framework.

This review has demonstrated the QOF’s limitations in improving health outcomes due to the indicators set and the method adopted in rewarding the QOF scores. It also highlights the need for further quality research, both on the long-term impact of the QOF on health outcomes and the effect of amending the current indicators set. This is essential to ensure that one of the largest pay-for-performance schemes in health is clinically and cost-effective in delivering quality care, which leads to optimal health outcomes.
Supplementary data

Supplementary data are available at the Journal of Public Health online.

References


2 NICE set to revise quality and outcomes framework indicators. Lancet 2009;374:588 (no authors listed).


