

## ANALYSIS



# Quality and Outcomes Framework: what have we learnt?



**Martin Roland** and **Bruce Guthrie** assess the successes and failures of the pay-for-performance scheme and what its future should be

Martin Roland *professor of health services research*<sup>1</sup>, Bruce Guthrie *professor of primary care medicine*<sup>2</sup>

<sup>1</sup>Institute of Public Health, Cambridge CB2 0SR, UK; <sup>2</sup>Population Health Sciences, University of Dundee, Dundee, UK

In 2004 the UK National Health Service introduced the largest health related pay-for-performance scheme in the world—the Quality and Outcomes Framework (QOF).<sup>1</sup> However Scotland is now abandoning the scheme, and growing disenchantment in England is likely to lead to major changes. What have we learnt, and what should happen to QOF in future?

## Promising start

In the late 1990s, general practitioners' pay had fallen substantially behind that of specialists, and morale and recruitment in general practice were poor. The government and the British Medical Association (BMA) privately agreed that a large pay rise was needed. Money was available because in 2000 the government had committed to increasing NHS spending to mid-European levels as a percentage of gross domestic product. However, the profession had to give something in return, and the BMA dropped its longstanding opposition to “quality payments” and started to negotiate a pay-for-performance scheme that would substantially increase funding for general practice.

There followed 18 months of negotiations between BMA and NHS Employers with a small number of clinical advisers to develop the outcome measures (indicators) that would form the basis of the scheme (box 1). A starting premise was that the clinical indicators should be based on evidence based guidelines so that they would be likely to command a wide degree of professional support (box 2). The framework also included indicators related to practice organisation and patient participation. The package was controversial, and the BMA allowed its members to vote on the scheme—once in outline and once when the details were known.

Implementation happened over more than a year. The indicators were available well before the financial rewards were introduced in order to facilitate planning, investment in electronic clinical records with tools for managing chronic disease, the production

of detailed guidance for practices, and support for implementation in some parts of the country. Electronic clinical records, which were already well advanced in primary care, became universal because they were needed to obtain payment, though GPs had to employ more administrative staff to collect the required data. QOF accelerated existing trends to shift care for chronic physical conditions to nurse-led clinics, particularly diabetes and cardiovascular and respiratory disease. Practices used the software tools created for QOF payments to monitor their care of patients, with more internal management in practices to ensure they met the targets.<sup>2</sup>

## High payments

With an overall cost of over £1bn (€1.2bn; \$1.3bn), QOF proved nearly £300m more expensive than the government had expected in the first year because it underestimated the baseline quality of care. This meant that many practices achieved near maximum performance (and therefore payment) in the first year. Practice income rose rapidly with QOF potentially providing an additional 25% of income, and this certainly reduced professional opposition to the scheme. However, initial rises in income were progressively clawed back over the next 10 years with zero or near zero pay rises such that real terms income in 2013-14 fell back to below that in 2003-04.<sup>3</sup>

All data from QOF were publicly available, and thus three major innovations were introduced simultaneously: much better data collection, public release of information on quality of care, and pay for performance. It is therefore unclear what effect the first two of these would have had on their own, and the degree to which pay for performance was a quality driver.

**Box 1: How the Quality and Outcomes Framework works**

- The original scheme included 76 clinical indicators covering 10 conditions
- Data on clinical quality were extracted automatically from practice electronic records
- Doctors could exclude patients from individual clinical indicators (exception reporting) for specified reasons including clinical inappropriateness, intolerance of medication, and patient dissent
- Organisational indicators included medical records, information for patients, education and training, practice management, and medicines management
- Patient experience indicators related to conducting and acting on the results of patient experience surveys and offering booked appointments of at least 10 minutes

**Box 2: Examples of indicators that attracted broad professional support**

- Percentage of patients aged  $\geq 45$  who have a record of blood pressure in the preceding five years
- Percentage of patients with coronary heart disease in whom the last blood pressure reading (measured in the preceding 12 months) is  $\leq 150/90$  mm Hg
- Percentage of diabetic patients with up-to-date influenza immunisation

The government decided that it would be discriminatory to put age limits on the indicators even though most of the available evidence was based on trials that excluded older people. As time went on, and with targets having largely been met, single disease indicators appeared less relevant to the needs of patients, particularly older people with multiple complex problems.

**Did QOF improve quality of care?**

QOF did produce some improvements in quality of care, but this was against a background of a widespread programme of quality improvement in the NHS that included national standards for the major chronic diseases, annual appraisal of all doctors working in the NHS, and widespread use of clinical audits to compare practices, sometimes with public release of data. So for asthma and diabetes, for example, the introduction of QOF was followed by a modest increase in the rate at which care was already improving.<sup>4</sup> For the major chronic diseases in QOF, there were also reductions in inequalities in delivery of care, with practices in socioeconomically deprived areas rapidly catching up with the performance of practices in more affluent areas.<sup>5</sup> The scheme may have limited the rise in emergency admissions for included conditions, but did not appear to reduce associated mortality.<sup>6-8</sup>

Quality of care for conditions that were not included also continued to improve, but at a slower rate than before the introduction of QOF.<sup>9</sup> There were almost certainly negative consequences—for example, the progressive decline in the ability of patients to see a GP of their choice was probably partly due to a relentless focus by government on incentives for rapid access to care.

**Did doctors cheat?**

Gaming and manipulation of data are hard to detect, and the planned “light touch” inspections were in practice lax. The government’s concern had been about one aspect of the scheme—the ability of doctors to exclude individual patients from the data (exception reporting) for a range of reasons, including their clinical judgment. This had been important to get professional support for the scheme, but the government saw it as an open invitation to game the system. This proved not to be the case, with only around 5% of patients reported as exceptions, though, as would be expected, rates of exception reporting were lower for simple processes such as measuring blood pressure and higher for more complex processes such as diagnosis and intermediate outcomes.<sup>10 11</sup>

**What went wrong with QOF?**

QOF remains one of the largest implementations of healthcare pay-for-performance in the world, and any programme on this

scale will experience difficulties. Over the years, there have been several technical problems, one of which is that the original payment formula unintentionally led to larger and more affluent practices getting systematically higher payments than smaller practices for the same level of quality.<sup>12</sup> A second major technical flaw was that payments based on responses to a national patient survey were subject to random variation such that practices could improve care from one year to the next but actually receive less money.<sup>13</sup>

Problems also occurred with some indicators after implementation. For example, the codes used to define diabetes registers were changed to include only records that stated the type of diabetes. Although intended to improve the quality of registers, some people with less specific diabetes codes effectively vanished from practices’ QOF registers, and these people may subsequently have received worse care.<sup>14</sup> Many of these problems could have been avoided by better testing of indicators before implementation, which eventually happened when the National Institute for Health and Care Excellence (NICE) took over development of indicators in 2009.

Although the initial indicators largely related to aspects of care that GPs already thought important, the alignment of indicators with professional values reduced over time. This was partly because the easy targets had already been met and new indicators were introduced that were evidence based but had only marginal gains despite high workload. In addition, an increasing proportion of QOF was taken up with indicators that met a managerial or policy agenda rather than a clinical one (table). There was also concern that the needs of the increasing population of older people with multiple complex problems were poorly served by indicators that focused exclusively on single diseases.

The maximum percentage of practice income linked to quality indicators was reduced from 25% to 15% in 2013 because of perceptions that the higher rate distorted clinical practice. More radically, GPs in one English district (Somerset) negotiated a complete alternative to QOF in 2015 and Scotland dropped QOF in 2016 in favour of a quality improvement scheme based on local “quality circles.”

The unpopularity of QOF among professionals has undoubtedly been increased by the administrative burden it produces at a time when GP workloads have been increasing, general practice has been receiving a declining share of the NHS budget, and work stress is higher than at any time in the last 15 years.<sup>15</sup>

## So what should happen now?

There is general agreement that it is a professional responsibility to maintain and improve quality of care but less agreement on how this should be done. Although there is consensus that QOF requires substantial change, evidence is conflicting on whether quality declines when pay-for-performance incentives are removed.<sup>16 17</sup> It would therefore be prudent to require some limited ongoing data collection to avoid serious adverse consequences of withdrawing financial incentives. There is also little evidence on what would work better, although the steady improvement in quality of care in the decade preceding QOF suggests that the development and implementation of guidelines and standards encouraged by local clinical audit is effective. The NHS in Scotland has chosen this approach, replacing QOF with “quality circles” implemented through clusters of 10-15 practices working collaboratively to identify and develop relevant improvement work. There is funding to release GPs from practices but no centrally created targets or financial incentives. Similar work is ongoing in Wales.

The successes of QOF included an acceleration of previous trends towards systematic management of chronic disease by multidisciplinary teams and widespread introduction of electronic medical records. However, quality and safety improvement require multiple strategies, sustained over time. Winning hearts and minds through persuasion, collaboration, and close alignment of professional and managerial agendas is at least as important as the more technical elements of any individual quality improvement initiative. QOF (and pay for performance more generally) was not a magic bullet to improve quality and reduce variation, but neither will its replacements be. It remains to be seen which of the divergent approaches being taken by the NHS in England, Scotland, Wales, and Northern Ireland is most successful.

**Contributors and sources:** Both authors have provided advice on the development of QOF and both contributed to the development and authorship of this article.

We have read and understood BMJ policy on declaration of interests and declare the following interests: MR advised the BMA and NHS Employers on the development of the Quality and Outcomes Framework from 2001 to 2003. BG was a member of the NICE Quality and Outcomes Framework indicators advisory committee and chaired the methods, retirement thresholds, and review subcommittee from 2009 to 2014.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 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. doi:10.1056/NEJMhpr041294pmid:15459308.
- McDonald R, Harrison S, Checkland K, Campbell SM, Roland M. Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study. *BMJ* 2007;334:1357-9. doi:10.1136/bmj.39238.890810.BEpmid:17580318.
- Health and Social Care Information Centre. GP earnings and expenses time series 2002-2014. 2015. <http://www.hscic.gov.uk/searchcatalogue?productid=13317&q=title%3a%22GP+Earnings+and+Expenses%22&sort=Relevance&size=10&page=1#top>
- Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med* 2009;361:368-78. doi:10.1056/NEJMsa0807651pmid:19625717.
- 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. *Lancet* 2008;372:728-36. doi:10.1016/S0140-6736(08)61123-Xpmid:18701159.
- Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ* 2014;349:g6423. doi:10.1136/bmj.g6423pmid:25389120.
- 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. *Lancet* 2016;388:268-74. doi:10.1016/S0140-6736(16)00276-2pmid:27207746.
- Roland M. Does pay-for-performance in primary care save lives? *Lancet* 2016;388:217-8. doi:10.1016/S0140-6736(16)00550-Xpmid:27207745.
- Doran T, Kontopantelis E, Valderas J, et al. Effect of financial incentives on incentivized and non-incentivized clinical activities. Evidence from the UK's Quality and Outcomes Framework. *BMJ* 2011;342:d3590. doi:10.1136/bmj.d3590pmid:21712336.
- Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med* 2008;359:274-84. doi:10.1056/NEJMsa0800310pmid:18635432.
- Roland M. Should doctors be able to exclude patients from pay-for-performance schemes? *BMJ Qual Saf* 2015;30:bmjqs-2015-005003. doi:10.1136/bmjqs-2015-005003pmid:26717988.
- Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. *Br J Gen Pract* 2006;56:836-41.pmid:17132350.
- Roland M, Elliott M, Lyraztopoulos G, et al. Reliability of patient responses in pay for performance schemes: analysis of national General Practitioner Patient Survey data in England. *BMJ* 2009;339:b3851. doi:10.1136/bmj.b3851pmid:19808811.
- 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. *BMJ* 2009;338:b1870. doi:10.1136/bmj.b1870pmid:19474024.
- Gibson J, Checkland K, Coleman A, et al. Eighth national GP worklife survey. University of Manchester, 2015. <http://www.population-health.manchester.ac.uk/healthconomics/research/Reports/EighthNationalGPWorklifeSurveyreport/EighthNationalGPWorklifeSurveyreport.pdf>
- Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer prescribing—a trial of education, informatics, and financial incentives. *N Engl J Med* 2016;374:1053-64. doi:10.1056/NEJMsa1508955pmid:26981935.
- Lester H, Schmittiel J, Selby J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ* 2010;340:c1898. doi:10.1136/bmj.c1898pmid:20460330.

Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to <http://group.bmj.com/group/rights-licensing/permissions>

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

**Key messages**

The Quality and Outcomes Framework accelerated previous trends towards widespread use of electronic medical records and multidisciplinary management of chronic diseases

QOF resulted in relatively limited additional improvements in quality but reduced socioeconomic inequalities in delivery of care

Several indicators were withdrawn because they lacked professional support or there were problems with implementation

New strategies are needed to continue improvements in quality of care

**Table****Table 1 | Examples of indicators that went wrong**

Indicator	Problem
Patient survey reports of whether patients could get appointments	Insufficient survey numbers and a poorly constructed formula linking survey scores to payment resulted in substantial random variation in payments (introduced in 2008, dropped in 2011)
Using a validated instrument (PHQ9) to assess the severity of depression within 28 days of a new diagnosis, repeating the assessment 2-12 weeks following diagnosis	Poor alignment with professional beliefs prompted substantial criticism about lack of supporting evidence (although other indicators with similar levels of evidence that GPs did believe in were not criticised in the same way). It was also easily "gamed" by GPs using free text description of the patient's problem rather than coding "depression" in electronic records (introduced in 2006, dropped in 2013)
Practices should develop a register of patients with obesity	Practices could effectively claim payments by including a register with one obese patient. The indicator does not encourage regular weighing to create a more comprehensive obesity register or any strategy for tackling the problems of obesity (introduced in 2006, still current)
Opportunistic screening of elderly and at-risk patients for dementia (technically an "enhanced service" rather than part of QOF)	Little professional support, substantial concern about harms resulting from false positive results, lack of services for specialist diagnosis and management (introduced in 2014, dropped in 2015)