HEDG Working Paper 08/21

Record rewards: the effect on risk factor monitoring of new financial incentives for UK general practices

Matt Sutton
Ross Elder
Bruce Guthrie
Graham Watt

August 2008

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

Record rewards: the effect on risk factor monitoring of new financial incentives for UK general practices

Running head: Effect of new incentives for UK general practices

Keywords: incentives, quality, primary care, payment systems, spillovers

Word count = 4,944. Table count = 6. Figure count = 1.

Matt Sutton, Health Economics Research Unit, University of Aberdeen

Ross Elder, Information Services Division, NHS National Services Scotland

Bruce Guthrie, Tayside Centre for General Practice, University of Dundee

Graham Watt, General Practice and Primary Care, University of Glasgow

Corresponding author: Matt Sutton, Health Economics Research Unit, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD. Tel: 01224 553733. Fax: 01224 550926. Email: m.sutton@abdn.ac.uk

Funding sources: This work was funded by NHS Quality Improvement Scotland as part of a project to examine clinical effectiveness using SPICE data. HERU receives funding from the Chief Scientist Office of the Scottish Government Health Directorate-General.

Conflicts of interest: None.

Abstract

An innovative and expensive performance-related pay scheme was introduced for general practices across the UK in 2004. It was not piloted and baseline performance data were not collected prior to its introduction. We estimate the impact of this Quality and Outcomes Framework (QOF) by analysing annual rates of recording of blood pressure, smoking status, cholesterol, body mass index and alcohol consumption based on individual patient records from 315 general practices over the period 2000/1 to 2005/6. The recording of each risk factor is designated as incentivised or unincentivised for each individual based on whether they have one of the diagnoses targeted by the QOF. The estimated impact is sensitive to the dynamic specification of the recording process and was substantially larger on the targeted patient groups (+19.9 percentage points) than the untargeted groups (+5.3). We also find positive spillovers of (+10.9) for the targeted groups onto unincentivised factors. We propose that the intended rewards per additional record were under-estimated, because account was not taken of substantial multiple-payment for co-morbid patients, levels of pre-QOF recording and the additional rewards available for risk factor control that would be achieved by measurement alone. Based on naïve assumptions, we estimate the intended financial reward per additional risk factor record to be £4.40. Allowing for co-morbidity, pre-QOF performance and the additional 'control' rewards, increases this average reward eleven-fold, to £48.90. Taking account of the positive spillovers reduces this figure to £25.10, but it remains substantially larger than what appears to have been intended.

[Word count = 248]

Introduction

The new contract for UK general practices introduced in April 2004 included a new funding stream for quality and outcomes, rewarding achievements on a wide range of quality indicators for ten targeted chronic diseases (NHS Confederation and British Medical Association, 2003). Introduction of this Quality and Outcomes Framework (QOF) represents a major initiative for improving quality in primary medical care (Roland, 2004).

The QOF is an expensive, innovative pay-for-performance scheme. Payments made under the QOF in Scotland totalled £134.5 million in 2005/06, approximately £27 per capita. It introduced a new data system for measuring quality and was introduced universally and simultaneously across the UK without any piloting. Average performance in the first year was very high but, since comparable data were not routinely collected prior to introduction, it is difficult to estimate the impact on quality of the introduction of this scheme.

There have been a number of studies that have compared quality before and after the introduction of the QOF. Gulliford et al (2007) and Tahrani et al (2007) concentrate on trends in incentivised indicators for patients with diabetes, and both find substantial increases in quality when the QOF was introduced. McGovern et al (2008) concentrate on trends in incentivised indicators for patients with coronary heart disease and find a similar effect. Hippisley-Cox et al (2007) examined trends over 2001-2006 in 19 incentivised indicators. Without presenting any formal analysis, they conclude that "[w]hilst there have been substantial increases in achievement of indicators since introduction of the new QOF in April 20004 (sic), there is good evidence that the changes predated the QOF give (sic) the increase observed since April 2001."

Campbell et al (2007) examined the quality of care for three of the conditions targeted by the QOF (CHD, asthma and diabetes). They compared trends in incentivised and unincentivised indicators for these patients with fitted trends. The 2003 to 2005 increases were significantly higher than the fitted trends for asthma and diabetes but only marginally so for CHD ($p=0.07$). The rate of improvement for the incentivised indicators did not differ significantly from the rate of improvement for the unincentivised indicators.

The study by Steel et al (2007) is the only one to consider unincentivised and incentivised indicators for targeted conditions (hypertension and asthma) and indicators for two untargeted conditions (osteoarthritis and depression). There were significant increases for both the incentivised and unincentivised indicators for the targeted conditions. The indicators for the untargeted conditions did not increase significantly.

None of the previous studies have exploited cleanly the presence of condition and risk factor combinations that are or are not incentivised. No previous study has compared the effects with the costs of the scheme. Moreover, no account has been taken of co-morbidity which, as we show, is important for assigning activities applied to individual patients to incentivised and unincentivised groups and for estimating the payment per additional record.

The Scottish Programme for Improving Clinical Effectiveness in Primary Care (SPICE-PC) is a quality initiative under the auspices of the Royal College of General Practitioners. A series of Care Management Screens was developed to embed national clinical guideline recommendations into data entry templates for electronic records. Practice participation is voluntary and involves confidential feedback on performance relative to the national average.

The new contract was introduced during the period in which SPICE has operated, offering the opportunity for before and after comparison of performance. Since SPICE collects information on indicators not measured and rewarded in the QOF, it also offers opportunity to assess whether quality has improved faster for incentivised activities.

In this paper, we use these data to analyse changes in quality over the period 2000/1 to 2005/6. We examine the recording of five risk factors for six patient groups. Unique among recent studies, we consider the entire population including a group of patients not diagnosed with any of five of the ten QOF conditions. We use a hierarchy of diagnoses to ensure that recording of some of the risk factors is incentivised for some patient groups and not incentivised for others.

Among the unincentivised risk factors we include body mass index (BMI), which is attracting growing attention as evidenced by inclusion of an obesity register in the post-2006 revision of the QOF (British Medical Association and NHS Employers, 2005), and alcohol consumption, which is emerging as a major cause of death (Leyland et al, 2007). We use these risk factors as a comparison group for the incentivised indicators and explicitly model the possibility of spillovers to these unincentivised activities for the targeted patients.

The QOF is a living incentive scheme. The included indicators, the points available and the upper thresholds at which maximum rewards are earned are under regular review. They were changed substantially from April 2006 (British Medical Association and NHS Employers, 2005). Since we do not adopt the QOF definitions of achievement, we are able to examine aspects of the design properties of the QOF including the level of payment, the upper thresholds and the period within which risk factors are required to be measured.

We find that the existence of incentives impacts directly on performance and has positive spillovers for targeted patients, and that the design of the incentives matters. We use these results to estimate the level of payment made to an average practice for each additional risk factor record. We derive figures for the intended payment per additional record based on naïve assumptions about levels of pre-QOF recording and no multiple-payment for recording of a single risk factor in patients with multiple targeted conditions. We then demonstrate the impact on these average reward figures of allowing for: co-morbidity; the observed changes in recording; additional rewards paid for 'control' of risk factors that are earned by measurement alone; and positive spillovers for targeted patients. We find that the combined effect of these factors increases the reward per additional record approximately six-fold.

Methods

The QOF incentive system

In the first two years of the QOF scheme, which we study here, practices were rewarded according to the performance they reported on 146 indicators. Through their performance on these indicators, practices earned up to 1,050 QOF points. Practices were rewarded for these points according to a complex, non-linear function of the prevalence of disease in, and size of, their registered populations (Guthrie et al, 2006; Gravelle et al, 2007). An average practice was paid £75 per point in the first year and £125 per point in the second year of the scheme.

Seventy-six of the indicators, and 57% of the financial rewards, were offered for the quality of clinical care. These 76 indicators were focused on the care of people with 10 targeted

chronic conditions and involved the maintenance of disease registers, the verification of diagnosis, the recording and management of risk factors, and the provision of selected treatments.

Achievement for each disease is assessed and paid for separately in the QOF. Consequently, there are significant economies of scope. The same individual patient will contribute to the rewarded performance for the practice across all indicators within the disease domain and, if the individual has more than one of the targeted diagnoses, the achievement of a particular indicator will contribute in each disease domain.

Points were awarded based on reported coverage rates if they were above a lower threshold of 25%. Maximum points were awarded if the coverage rate exceeded an upper threshold, which varied across indicators. Practices were allowed to remove patients from the denominator used to calculate the coverage rate if treatment was inappropriate or refused. This potential for 'exception reporting' of patients, and the existence of the upper thresholds, have been shown to influence practice behaviour (Gravelle et al, 2007).

In this paper we analyse the recording of five risk factors. Table I provides information on whether the recording of each of these five risk factors was incentivised in the QOF. Where included, the maximum points available for recording this risk factor, the time period in which a record is required to qualify as 'achievement', and the upper threshold required for maximum attainment, are shown. In addition to the targeted diagnosis groups, practices could earn points for "records and information about patients", which includes up to 11 points for recording the smoking status of all patients aged 15–75 years and up to 15 points for recording the blood pressure of all patients aged 45 years and over in the preceding five years.

The upper threshold is 90% for most indicators with the exception of the two indicators for all patients, which reach a maximum at 75%. For most indicators, achievement for an individual patient requires that the risk factor is recorded within the last fifteen months. Blood pressure recording for hypertension patients is required more frequently - every nine months. Recording of blood pressure for untargeted patients is required less frequently – within the last five years. Recording of smoking status for hypertensive patients is only required once since diagnosis, and for untargeted patients only once ever.

In the Care Management Screens created by SPICE, recording of information is recommended for more risk factors than are rewarded on the QOF for several conditions. We define these as other 'clinically-effective' activities. Table I indicates instances where risk factor recording is included in the SPICE criteria but not rewarded on the QOF. For diabetes and CHD, collection of information on all risk factors is recommended. For hypertension, all risk factors except cholesterol are recommended for periodic collection. The COPD Screen did not exist before the QOF was introduced.

Some QOF activities facilitated access to additional quality points. Once blood pressure and cholesterol had been recorded, practices could earn additional points if the last recorded measurement was 'controlled' (i.e. under a specified value). Practices could also earn additional points for recording that they had offered cessation advice to patients whose current smoking status had been established. The potential of these additional rewards offered further incentives to practices to increase the recording of patient risk factors.

Data

The SPICE data were provided by the Primary Care Clinical Informatics Unit, Department of General Practice, Aberdeen University. A full extract of the data up to and including the Spring quarter of 2006 containing data from 315 general practices was imported into the NHSScotland Information Services Division.

We considered five chronic conditions included in the QOF: Coronary Heart Disease (CHD); Chronic Obstructive Pulmonary Disease (COPD); Diabetes; Hypertension; and Stroke. Individuals were considered diagnosed if they had received a relevant diagnosis (using the Read system) prior to the start of the financial year. Details of the Read codes used are in the full report (Elder et al, 2007). The accuracy of these diagnosis codes was assessed by comparisons with the disease register sizes reported by practices for the QOF in 2005/6 (Elder et al, 2007).

Participation in SPICE is voluntary and information may be only partially recorded. To be suitable for inclusion, the indicators should have reasonable coverage prior to the introduction of the QOF. Rates of recording of five risk factors were selected as they were consistently collected and were of relevance across a range of conditions. The five risk factors were: smoking status; alcohol consumption; blood pressure; cholesterol; and BMI.

The analysis is restricted to patients aged 45 years and over. Although this is only 37.1% of the full dataset, it represents 92.2% of the targeted diagnosis groups and is the age cut-off for the requirement to record blood pressure for all patients under the QOF. We make this restriction to focus on the population for whom risk factor recording is most important.

Representativeness of participating practices

A range of practice characteristics for the 315 practices in the sample were compared to the 721 other Scottish practices (Elder et al, 2007). Participation in SPICE was less likely in the most deprived areas and showed some geographical concentration. Participating practices had more patients in total but fewer patients per GP than non-participants, were more likely to also participate in other voluntary initiatives and achieved 1% points more on average on the 2005/6 QOF. This suggests some caution in extrapolating the results to all Scottish practices. However the differences on each variable are relatively small and, since characteristics could not be attached to individual practices to ensure data confidentiality, we are unable to correct for selection bias.

Mutually exclusive diagnosis groups

Individual patients can appear in more than one diagnostic group. The designation of indicators as incentivised and unincentivised requires knowledge of all the diagnoses received by individuals and definition of mutually exclusive diagnosis groups. We therefore created a hierarchy of diagnoses based on the number of risk factors incentivised in the QOF.

Each individual patient appears in just one of the following six groups:

1. *Diabetes*
2. *CHD* (excluding the above)
3. *Stroke* (excluding all above)
4. *Hypertension* (excluding all above)
5. *COPD* (excluding all above)

6. *Untargeted*

There is considerable overlap in these diagnoses. The percentages by which each group is reduced when we allow for co-morbidity are: CHD=18%; Stroke=38%; Hypertension=34%; COPD=54%. It is important, therefore, to take account of co-morbidity in assigning individual patients to incentivised/unincentivised categories on each risk factor.

The other conditions included in the QOF in 2004/5 and 2005/6 were asthma, epilepsy, cancer, mental health and hypothyroidism. Only in the case of asthma are any of the risk factors we study incentivised and that is only smoking status. We were unable to isolate patients included in the asthma domain from the diagnostic information because this also requires data on prescription of asthma-related drugs. The published statistics (www.isdscotland.org/QOF) indicate that 5.3% of the Scottish population were included in the asthma domain of the QOF, though almost two-thirds are under 45 years-old and are excluded from this analysis.

Analysis

The unit of analysis is each risk factor for each patient in each year. Within a year therefore, there are five observations for each patient. The dataset is unbalanced over time because patients are included only if they have been registered throughout that year. Patients that are registered with a practice throughout all five years appear 30 (5 risk factors x 6 observation years) times. Alongside each risk factor record, we have an indicator of the diagnosis group, the practice, the year and the patient's age and sex. To this dataset, we match on the details of the incentive scheme shown in Table I.

The specification of the basic model is:

$$y^*_{ijkt} = x'\alpha + \beta_j + \gamma_k + \delta_t + \theta q_{jk} + \varepsilon_{ijkt} \quad (1)$$

in which y^*_{ijkt} is the latent quality index for individual *i* on risk factor *j* in condition group *k* at time *t*. *x* is a vector of interactions between gender and a cubic function of age. The parameters β, γ and δ represent fixed effects for the risk factors, condition groups and years respectively. The binary variable q_{jk} indicates whether the recording of risk factor *j* is incentivised by the QOF for condition *k*.

Equation (1) assumes that pre-QOF recording of risk factors is not influenced by whether particular diagnosis-indicator combinations will become incentivised in the QOF. Since the incentivised diagnosis-indicator combinations were selected on the basis of evidence of clinical effectiveness (Roland, 2004), we expect their recording rates to be higher pre-QOF. To allow for this, we include an interaction between q_{jk} and a dummy variable (*D*) indicating the post-QOF period.

$$y^*_{ijkt} = x'\alpha + \beta_j + \gamma_k + \delta_t + \theta_1 q_{jk} + \theta_2 D[t >= 4] q_{jk} + \varepsilon_{ijkt} \quad (2)$$

The parameter θ_2 measures the effect of the introduction of the QOF on the specific diagnosis-indicator combinations that become incentivised, relative to their pre-QOF levels. The parameter θ_1 measures the extent to which these combinations were recorded at different rates to other combinations pre-QOF. θ_2 is the incremental effect of the QOF on incentivised activities, conditional on the general time trend.

Equation (2) assumes that there are no spillovers of the introduction of the QOF into unincentivised diagnosis-indicator combinations. There may be spillovers onto other indicators for targeted diagnoses. We test this by including an interaction between γ_k and the dummy variable for the post-QOF period. The coefficient on this interaction indicates whether there is a change in the recording of unincentivised risk factors for patient groups targeted by the QOF. To examine whether this spillover is onto clinically-effective activity, we separately model the effects on those indicators included and those indicators not included in the SPICE Care Management Screens.

Finally we analyse the design properties of the QOF indicators by considering an extension to equation (2) in which we expand the specification of q_{jk}. We replace q_{jk} with:

- the financial reward available for maximum achievement on this indicator divided by the number of patients in this diagnosis group for an average practice;
- the upper threshold at which maximum achievement is awarded for this indicator for this diagnosis group; and
- the width of the period (in years) in which the QOF requires recording of this risk factor to be considered 'achieved'.

Each of these variables is included as a main effect and as an interaction with the dummy variable indicating the post-QOF period (D). As in equation (2) the coefficients on the interaction terms capture differential changes in recording following introduction of the QOF.

Econometric analysis

The latent index y^*_{ijkt} is not observed. We assume a probit link function to the binary observed indicator of whether the risk factor has been recorded and allow for unobserved individual heterogeneity using a random-effects specification. In the tables we present coefficients but calculate average partial effects for the effect of the incentives by comparing individual predictions in the post-QOF period with predictions setting *D* to zero.

We analyse the sensitivity of the results to inclusion of variables reflecting the dynamic structure of the process. We include the lag of the dependent variable ($y_{ijk,t-1}$) and the initial value ($y_{ijk,t_{I\min}}$) observed for the individual (Contoyannis et al, 2004). These variables are designed to model a stable dynamic process. Such dynamics may be interrupted if the individual changes the practice with which they are registered and receives a comprehensive health check including the collection of some of the risk factors we consider. To allow for this possibility we include an interaction between $y_{ijk,t_{I\min}}$ and whether the individual was a new patient when they are first observed (t_{Imin}).

We also include a binary variable indicating whether the individual will drop out of the sample in the next year. This will occur if the individual de-registers from the practice. There are known delays in the de-registration process so we might anticipate that the individual will be less likely to have had their risk factors recorded in the year prior to the date they are officially recorded as no longer registered.

Estimation of the rewards per record

An expression for the cost per additional risk factor record (*AC*) to be supplied by an average practice that was designed into the incentive system is:

$$AC = \frac{\lambda \sum_{jk} \pi_{jk}}{\sum_{jk} N_{jk}} = \frac{\lambda \sum_{jk} \pi_{jk}}{\sum_{jk} (M.p_k(u_{jk} - l_{jk}) f_{jk}^{-1})} \quad , 0 \le p_k \le 1, 1 \le \sum_k p_k \le K \qquad (3)$$

in which λ is the reward per point (£125 in 2005/6), π_{jk} is the maximum available points for recording risk factor *j* on diagnosis group *k*, and N_{jk} is the number of risk factor records required to achieve the maximum points. Each N_{jk} is given by the product of: the practice population (*M*); the prevalence rate (*p*) for diagnosis *k*; the difference between the upper and lower thresholds; and the inverse of the period in years with which the risk factor must be measured (f_{jk}). Note that the sum of the prevalence rates can exceed one because, for this naïve estimate, individual patients can appear in more than one diagnosis group.

Equation (3) over-estimates the number of records required because it does not allow for the possibility that individual patients can appear in more than one diagnostic group. Further, it assumes that pre-QOF levels of recording were at the lower thresholds. Finally, Equation (3) only includes the direct financial rewards from risk factor recording. To include the rewards that a practice will receive for recording a risk factor and discovering that it is under the critical value required to qualify for the 'control' points, we must add a proportion of these additional points.

A measure of the actual rewards per record (*AC'*) is given by:

$$AC' = \frac{\lambda\left(\sum_{jk}\pi_{jk} + \sum_{jk}\eta_{jk}\pi'_{jk}\right)}{\sum_{jk}(P_k.APE(\hat{\theta}_2))} \quad , \sum_k P_k = M \qquad (4)$$

in which π'_{jk} is the additional points available for controlling the risk factor and η_{jk} is the proportion of the points available for risk factor control that are earned through records that are 'controlled' when first measured and require no additional effort for achievement. The P_k are prevalence counts for the K mutually exclusive and exhaustive groups. The estimated additional records generated are obtained by multiplying by the average partial effect ($APE(\hat{\theta}_2)$).

All of these parameters are specified by the payment system with the exception of p_k, P_k, θ_2 and η_{jk}. We derive estimates of AC and AC' using prevalence estimates from the published QOF data (www.isdscotland.org/QOF) and estimates of the extent of co-morbidity in our dataset. We use the proportions with risk factors already controlled upon diagnosis published in our full report on the SPICE data (Elder et al, 2007, Table 5.9) to estimate the number of points that the average practice will receive for risk factor 'control' by recording alone. We do not include any of the additional points available for smoking cessation advice since they require some additional effort to secure achievement for the current smokers.

Finally, we examine how the estimate of AC' changes when we allow for the records generated as positive spillovers from the introduction of the QOF. This increases the denominator while having no impact on the numerator.

Results

Summary statistics

Summary statistics for this dataset of 9.4million observations are provided in Table II. Risk factors are recorded on one quarter of occasions. The mean age of patients is 61 years and 47% are male. The proportions with the targeted diagnoses (excluding co-morbidity) are: 6.6% for diabetes, 9.0% for CHD, 15.6% for hypertension, 1.7% for COPD and 2.7% for stroke. In the dataset 45% of observations are for combinations of risk factors and conditions that become incentivised in the QOF.

The general trends in the rates of recording of risk factors are exemplified by Figure 1, which shows the annual rates for patients diagnosed with CHD (and not diabetes). Under the QOF, the recording of blood pressure, smoking status and cholesterol become incentivised for this group, while the recording of BMI and alcohol consumption does not. All risk factors show an increase over the period but there is a marked increase for smoking status and cholesterol, which become incentivised and which were recorded at similar rates to the unincentivised factors in the early years.

Dynamics

The impact of adding the dynamic specification into the regression model is demonstrated in Table III. Model (1) is the most basic specification, which contains demographic variables, dummy variables for year, diagnosis and risk factor, and a dummy for whether the diagnosis-factor combination is incentivised. Risk factor recording varies by sex and age, and increases

over the period 2000/1 to 2004/5. Risk factor recording is higher for all of the targeted conditions than for the rest of the population. The rates are highest for diabetes and CHD. Rates of recording are higher than the reference risk factor (alcohol consumption) for blood pressure, smoking status and BMI.

The coefficient on the dummy variable indicating that a diagnosis-factor is incentivised is reduced when a variable is introduced to reflect the higher rates of recording of the incentivised diagnosis-factor combinations prior to the introduction of the QOF (Model (2)). The effect is further reduced when the variables capturing the dynamic process are introduced (Model (3)). The average partial effect of the 'is incentivised' dummy variable in Model (3) equals 7.9 percentage points.

In Model (3) both the one-year lagged and first period values of the dependent variable exert a significant, positive effect on the probability of risk factor recording. As expected, the magnitude of the effect of the initial condition is reduced if it is measured on a new patient. Also as expected, individuals that will drop out of the sample in the next year have a lower probability of a risk factor record.

Spillovers for targeted patients

In Table IV we analyse whether the effects of the incentives differed between targeted and untargeted diagnostic groups and test for spillovers for the targeted patients. Model (1) separates the effects of the incentives between Group 0 (the targeted diagnoses) and Group 1 (the untargeted group). Rates of recording were higher in the pre-intervention period for both groups. Average partial effects indicate that, following the introduction of the incentives, risk

factor recording increased by 15.4 percentage points for the targeted group and by just 1.4 percentage points for the untargeted group.

Model (2) additionally includes the potential spillovers for the targeted patients. We present the effects for the clinically effective factors (Group 2) and the remaining factors (Group 3) separately. The recording rate for the clinically effective factors was higher than the other factors in the pre-intervention period. Average partial effect calculations indicate that, following the introduction of the QOF, recording of unincentivised, clinically effective factors for the targeted groups increased by 10.9 percentage points and by 5.0 percentage points for the remaining factors. Because they no longer contribute to the reference category against which the incentivised factors are compared, allowing for these spillovers increases the estimated average partial effects of the incentives for both the targeted and untargeted groups of patients to 19.9 and 5.3 percentage points respectively.

Sensitivity of the response to the design features of the incentives

In Table V we examine whether the responses to the incentives were sensitive to their design. Before the QOF is introduced, rates of recording are significantly higher for condition and risk factor combinations that will attract more reward, require higher rates to reach the upper threshold and where the QOF measurement period is wider. After the QOF is introduced, recording rates increase significantly more for those combinations that offer greater reward and require higher achievement to reach the upper threshold and reduce for those where a wider measurement period is allowed for qualifying for QOF achievement.

Estimated rewards per record

The average practice in Scotland has 5,333 patients. Our simplest estimate of the intended cost per record designed into the QOF scheme is £4.37, since an average practice achieving maximum points will receive £12,750 for 2,916 risk factor records (Table VI). This assumes that the scheme was intended to compensate practices for all risk factor records. Restricting the denominator to those above the lower thresholds (as in Equation (3)) increases the intended reward per additional record to £5.56. Allowing for individual patients that have more than one diagnosis increases the intended average reward per additional record to £7.78.

We have estimated that the average practice generated an additional 411 records for the incentivised diagnosis-factor combinations in response to the QOF and received 99% of the maximum payment (£12,610/£12,750). The actual rewards per additional record were therefore considerably higher than intended (£30.65). When we include the estimated additional rewards received by practices for records that qualified for achievement under the 'control' criteria without additional effort, the estimated reward per additional record increases to £49.86.

However, our analysis suggests that practices responded to the introduction of payments for targeted patients by increasing risk factor recording across the board for these patients. Allowing for these effects increases our estimate of the direct effect of the scheme (since these no longer contribute to the general trend) and adds an indirect, positive effect. When we include these positive spillovers for targeted patients, the estimated average reward per additional record is £26.26 (when we include clinically-effective spillovers only) or £25.06 (when we include all spillovers).

Discussion

The availability of SPICE data before the introduction of the QOF provides a rare opportunity to assess the contribution of the QOF to recent improvements in the quality of patient care in general practice. We focused on the recording of five risk factors across six mutually exclusive groups of patients. Five of these six patient groups were defined by QOF conditions and a hierarchy of diagnoses was defined to ensure that the recording of some risk factors was not incentivised in the QOF for some groups. The recording of one risk factor (alcohol consumption) was not incentivised for any of the groups despite being recommended on the SPICE Care Management Screens for three of the groups.

We found that rates of recording increased for all risk factors for all groups. Nevertheless, rates of recording increased most rapidly when it became explicitly incentivised in the QOF. We have estimated the overall increase in recording on the incentivised indicators to be 19.9 percentage points for targeted patients and 5.3 for untargeted patients. In addition, we have identified a 10.9 percentage point increase in recording of clinically-effective, unincentivised factors for targeted patients, which may represent positive spillovers. As we might expect, responses were greater on indicators that attracted more payment and required more stringent performance. However the more generous width of the measurement period required for achievement on some indicators appears to have led to reductions in their recording rates compared to other indicators.

The average cost figures based on the additional records directly generated are up to eleven times larger than the figures implicitly designed into the scheme. Practices were rewarded close to the maximum amounts available whilst generating only a seventh of the records

thought to be required as additional effort. Previously this effort was provided within the funding they received under a predominantly weighted capitation contract and these sums remained protected under the new contract. Including the positive spillovers as an effect of the incentive scheme reduces the average costs by a factor of two, but they still remain significantly above those implicitly designed into the scheme. This provides further evidence, if it were needed, that these expensive incentives should have been more carefully designed and piloted prior to introduction.

Acknowledgements

We are grateful to other members of the project team (Bill Ramsay, Michelle Fitzpatrick, Colin Fischbacher, Colin Simpson, Margaret MacLeod) and to participants in the summer 2007 Health Economists' Study Group meeting for comments on previous versions of this work.

References

British Medical Association and NHS Employers. 2005. *Revisions to the GMS contract 2006/07. Delivering investment in general practice*.

Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. 2007. Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine* **357(2)**: 181-90.

Contoyannis P, Jones AM, Rice N. 2004. The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics* **19**: 473-503.

Elder R, Kirkpatrick M, Ramsay W, Macleod M, Guthrie B, Sutton M, Watt G. 2007. *Measuring quality in primary medical services using data from SPICE*. Clinical Indicators Support Team, ISDScotland, Edinburgh. www.indicators.scot.nhs.uk

Gulliford MC, Ashworth M, Robotham D, Mohiddin A. 2007. Achievement of metabolic targets for diabetes by English primary care practices under a new system of incentives. *Diabetic Medicine* **24(5)**: 505-11.

Hippisley-Cox J, Vinogradova Y, Coupland C. 2007. *Time Series Analysis for selected clinical indicators from the Quality and Outcomes Framework 2001-2006*. Final Report for The Information Centre for health and social care. QRESEARCH and The Information Centre for health and social care.

Leyland AH, Dundas R, McLoone P, Boddy A. 2007. *Inequalities in mortality in Scotland 1981-2001*. MRC Social and Public Health Sciences Unit. Occasional Paper no.16.

McGovern MP, Boroujerdi MA, Taylor MW, Williams DJ, Hannaford PC, Lefevre KE, Simpson CS. 2008. The effect of the UK incentive-based contract on the management of patients with coronary heart disease in primary care. *Family Practice* forthcoming.

Gravelle H, Sutton M, Ma A. 2008. *Doctor Behaviour under a Pay for Performance Contract: Further Evidence from the Quality and Outcomes Framework.* Research Paper 34. Centre for Health Economics: University of York.

Guthrie B, McLean G, Sutton M. 2006. Workload and reward in the quality and outcomes framework of the 2004 general practice contract. *British Journal of General Practice* **56**: 836-41.

NHS Confederation and British Medical Association. 2003. *New GMS contract: Investing in General Practice*. London.

Roland M. 2004. Linking physician pay to quality of care: a major experiment in the United Kingdom. *New England Journal of Medicine* **351**:1448-54.

Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. 2007. Quality of clinical primary care and targeted incentive payments: an observational study. *British Journal of General Practice* **57**: 449-54.

Tahrani AA, McCarthy M, Godson J, Taylor S, Slater H, Capps N, Moulik P, Macleod AF. 2007. Diabetes care and the new GMS contract: the evidence for a whole county. *British Journal of General Practice* **57**: 483-5.

Table I: Design of the Quality and Outcomes Framework financial incentives

Diagnosis-factor	QOF			Not in QOF but designated clinically-effective*	Potential additional points facilitated†
	Points available	Measurement period	Upper threshold		
Diabetes					
- Blood pressure	3	Every 15 months	90%	-	17
- Body Mass Index	3	Every 15 months	90%	-	-
- Cholesterol	3	Every 15 months	90%	-	6
- Alcohol consumption	-	-	-	Yes	-
- Smoking status	3	Every 15 months	90%	-	5
CHD					
- Blood pressure	7	Every 15 months	90%	-	19
- Body Mass Index	-	-	-	Yes	-
- Cholesterol	7	Every 15 months	90%	-	16
- Alcohol consumption	-	-	-	Yes	-
- Smoking status	7	Every 15 months	90%	-	4
Stroke					
- Blood pressure	2	Every 15 months	90%	-	5
- Body Mass Index	-	-	-	-	-
- Cholesterol	2	Every 15 months	90%	-	5
- Alcohol consumption	-	-	-	-	-
- Smoking status	3	Every 15 months	90%	-	2
Hypertension					
- Blood pressure	20	Every 9 months	90%	-	56
- Body Mass Index	-	-	-	Yes	-
- Cholesterol	-	-	-	-	-
- Alcohol consumption	-	-	-	Yes	-
- Smoking status	10	Once since diagnosis	90%	-	10
COPD					
- Blood pressure	-	-	-	-	-
- Body Mass Index	-	-	-	-	-
- Cholesterol	-	-	-	-	-
- Alcohol consumption	-	-	-	-	-
- Smoking status	6	Every 15 months	90%	-	6
Untargeted					
- Blood pressure	15	Every 5 years	75%	-	-
- Body Mass Index	-	-	-	-	-
- Cholesterol	-	-	-	-	-
- Alcohol consumption	-	-	-	-	-
- Smoking status	11	Ever	75%	-	-

Notes: *Indicates that recording of this diagnosis-factor combination is recommended in the Care Management Screens designed by the Scottish Programme for Improving Clinical Effectiveness in Primary Care. † For those patients whose blood pressure and/or cholesterol are measured, additional points can be earned for reaching a level of control. For those patients whose notes record current smoking, additional points can be earned for offering cessation advice.

Table II Summary statistics

Variable	Mean	Minimum	Maximum
Male	0.473	0	1
Age (years/100)	0.610	0.45	1
New patient this year	0.114	0	1
Patient drops out next year	0.045	0	1
2000/1	0.134	0	1
2001/2	0.145	0	1
2002/3	0.158	0	1
2003/4	0.173	0	1
2004/5	0.189	0	1
2005/6	0.202	0	1
Diagnosis group	0.089		
CHD	0.017	0	1
COPD	0.066	0	1
Diabetes	0.156	0	1
Hypertension	0.027	0	1
Stroke	0.644	0	1
Untargeted		0	1
	0.246		
Risk factor recorded	0.450	0	1
Diagnosis-indicator combination becomes incentivised	0.363	0	1
	4.394		
Upper threshold (proportion)	1.857	0	0.90
Required measurement period (years)	0.473	0.75	5.00
Indicator payment per patient in average practice (£)	0.610	0	15.63

Notes: Figures based on 9,416,130 observations on 5 risk factors for 391,323 individuals in each of up to 6 years.

Table III Effects of dynamic specification on estimate of direct effect of incentives

Model	**(1)**		**(2)**		**(3)**	
Variable	**Coeff.**	**z**	**Coeff.**	**z**	**Coeff.**	**z**
Age (years/100)	-29.283	-29.0	-29.372	-29.0	-25.647	-25.7
Age^2	47.192	31.1	47.339	31.1	41.718	27.8
Age^3	-24.981	-33.6	-25.061	-33.6	-22.205	-30.2
Male	0.651	1.9	0.615	1.8	0.549	1.6
Male*Age	-7.330	-4.5	-7.197	-4.4	-6.645	-4.1
Male*Age^2	14.930	6.0	14.763	5.9	13.693	5.6
Male*Age^3	-8.274	-6.7	-8.204	-6.6	-7.604	-6.2
2002/3[+]	0.214	86.4	0.217	87.0	0.218	87.0
2003/4[+]	0.732	310.7	0.742	312.4	0.743	310.3
2004/5[+]	0.940	351.7	1.042	377.5	1.042	373.7
2005/6[+]	0.930	347.3	1.033	373.3	1.017	358.3
CHD[†]	1.475	300.5	1.368	275.6	1.291	263.8
COPD[†]	0.487	49.4	0.490	49.4	0.455	46.8
Diabetes[†]	2.041	372.5	1.814	322.6	1.674	298.1
Hypertension[†]	1.142	303.2	1.153	304.3	1.063	282.2
Stroke[†]	1.021	128.6	0.916	114.5	0.855	109.1
Blood pressure[‡]	0.869	369.6	0.248	60.7	0.189	46.0
Cholesterol[‡]	-0.087	-40.0	-0.273	-111.4	-0.259	-105.8
Smoking status[‡]	0.480	202.2	-0.132	-32.4	-0.138	-33.8
Body Mass Index[‡]	0.238	115.5	0.187	90.0	0.165	79.1
y_{t-1}					0.029	16.2
$y_{t=0}$					0.097	64.0
$y_{t=0}$*new patient					-0.106	-19.7
Dropout next year					-0.172	-39.9
Becomes incentivised			0.693	185.2	0.677	180.2
Is incentivised	**0.435**	**184.7**	**0.263**	**103.4**	**0.256**	**99.9**
Constant	3.642	16.6	3.639	16.5	2.826	13.0
Number (observations)	7088830		7088830		7088830	
Number (individuals)	391323		391323		391323	
Log-likelihood	-2871796		-2854451		-2840194	

[+] Reference category = 2001/2. [†] Reference category = 'untargeted' patients. [‡] Reference category = alcohol status.

Table IV Testing for spillovers

Model	**(1)**		**(2)**	
Variable	**dF/dx**	**z**	**dF/dx**	**z**
Group 0¶	0.589	147.5	0.766	89.4
Group 1¶	0.700	163.2	0.626	142.7
Group 2¶			0.113	13.6
Group 0¶ * post-QOF	**0.592**	**173.8**	**0.729**	**186.9**
Group 1¶ * post-QOF	**0.043**	**14.3**	**0.163**	**48.2**
Group 2¶ * post-QOF			**0.296**	**74.2**
Group 3¶ * post-QOF			**0.184**	**19.4**
Number (observations)	7088830		7088830	
Number (individuals)	391323		391323	
Log-likelihood	-2827733		-2824581	

Models also contain variables shown in top five panels of Table III. ¶ Reference category = 'unincentivised factors for untargeted patients'. Group 0 is 'incentivised factors for targeted patients'; Group 1 is 'incentivised factors for untargeted patients'; Group 2 is 'clinically effective unincentivised factors for targeted patients'; Group 3 is 'other unincentivised factors for targeted patients'.

Table V Modelling the effect of the design properties of the QOF

Variable	dF/dx	z
Indicator payment per patient	0.0067	13.0
Upper threshold	0.6042	119.1
Measurement period (years)	0.0082	5.4
Indicator payment per patient * post-QOF	**0.0265**	**34.8**
Upper threshold * post-QOF	**0.1292**	**33.5**
Measurement period (years) * post-QOF	**-0.0813**	**-40.8**
Number (observations)	7088830	
Number (individuals)	391323	
Log-likelihood	-2823216	

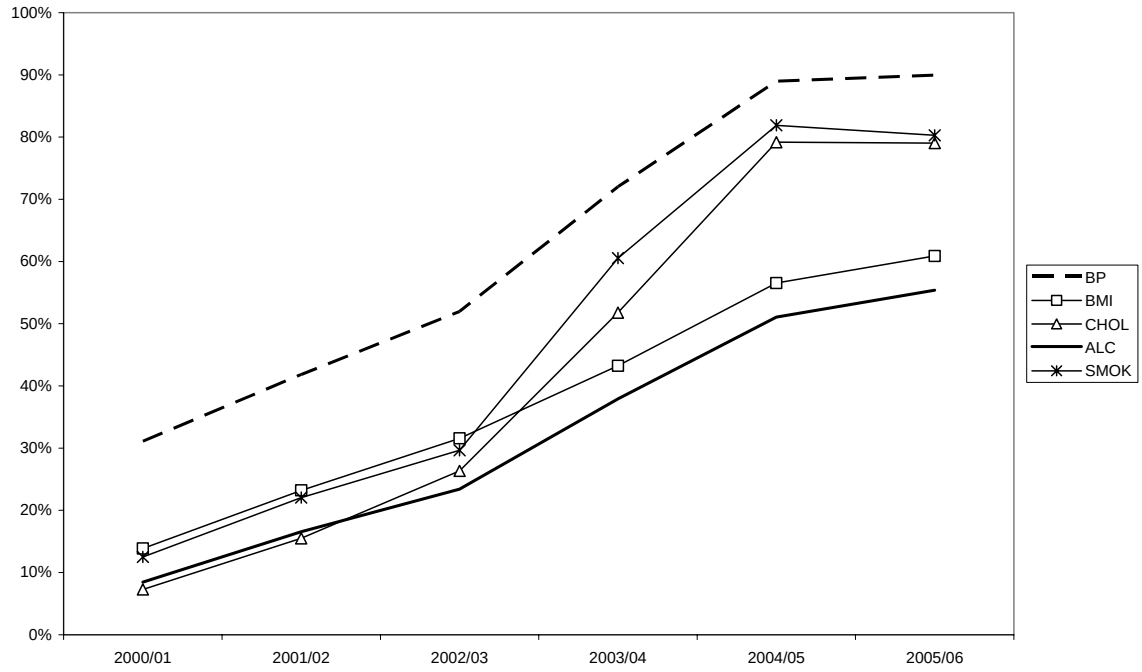
Models also contain variables shown in top five panels of Table III.

Table VI Financial rewards per risk factor record for an average practice under different scenarios

Intended/Actual	Basis of estimate of number of records	Allows for co-morbidity	Includes payment for 'control' [+]	Includes spillovers for targeted patients	Payment	Records	Payment per record
Intended	From zero to upper thresholds	No	No	No	£12,750	2,916	£4.37
Intended	From lower to upper thresholds	No	No	No	£12,750	2,294	£5.56
Intended	From lower to upper thresholds	Yes	No	No	£12,750	1,639	£7.78
Actual	Estimated (Model (1), Table IV)	Yes	No	No	£12,610	411	£30.65
Actual	Estimated (Model (1), Table IV)	Yes	Yes	No	£20,516	411	£49.86
Actual	Estimated (Model (2), Table IV)	Yes	Yes	Effective only[†]	£20,516	781	£26.26
Actual	Estimated (Model (2), Table IV)	Yes	Yes	All	£20,516	819	£25.06

[+] Achieved by recording alone. [†] Diagnosis-factor combinations recommended in the Care Management Screens designed by the Scottish Programme for Improving Clinical Effectiveness in Primary Care.

Figure 1 Annual rates of recording of five risk factors for patients diagnosed with Coronary Heart Disease



Notes: Excludes patients with diabetes. BP = Blood Pressure; CHOL = Cholesterol; SMOK = Smoking Status. Risk factors not incentivised: BMI = Body Mass Index; ALC = Alcohol Status. BP, CHOL and SMOK became incentivised for CHD in 2004/5; BMI and ALC were not incentivised for CHD.