Research Paper 12/02

The Tougher the Better: The Effect of an Increased Performance Threshold on the Performance of General Practitioners

Yan Feng, Ada Ma, Shelley Farrar and Matt Sutton

April 2012

The Tougher the Better: The Effect of an Increased Performance Threshold on the Performance of General Practitioners

Yan Feng[a], Ada Ma[b], Shelley Farrar[b], Matt Sutton[c]

[a]Office of Health Economics, London, UK

[b]Health Economics Research Unit, University of Aberdeen, Aberdeen, UK

[c]Health Economics, Health Sciences Research Group, University of Manchester, Manchester, UK

OHE Research Paper 12/02

April 2012

Corresponding author:

Yan Feng
Economist
Office of Health Economics
7th Floor, Southside, 105 Victoria Street
London SW1E 6QT
Email: yfeng@ohe.org
Phone +44 (0)207 747 8863

KEY WORDS: financial incentives, quality and outcomes framework, general practitioners

Acknowledgements

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication. Once a version of the Research Paper's content is published in a peer-reviewed journal, this typically supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

The research for this paper was carried out while Yan Feng was a PhD student in the Health EconomicsResearch Unit (HERU), University of Aberdeen, which is funded by the Chief Scientist Office of the Scottish Government Health Directorates (SGHD). The PhD was funded by the Institute of Applied Health Sciences University of Aberdeen PhD studentship. The current paper was written when Yan was at the Office of Health Economics (OHE) and was funded from the annual programme grant OHE receives from the ABPI.

We would like to thank Luís Silva Miguel and the audience for their comments on this work during the tenth European Health Economics Workshop. We also thank Hugh Gravelle, Nancy Devlin, Stephen Morris, Marjon van der Pol, Robert Elliot and Mandy Ryan for their comments and suggestions.

Any views expressed are those of the authors, and do not necessarily reflect the views or approval of OHE, its Editorial Board, its sponsors, HERU or SGHD.

Abstract

This paper investigates whether and how a change in the performance-related payment scheme motivated General Practitioners (GPs) in Scotland. It evaluates the effect of increases in April 2006 in the performance thresholds required for maximum payment under the Quality and Outcomes Framework (QOF). A difference-in-differences estimator with fixed effects was employed to examine the number of patients treated under clinical indicators whose payment schedules were revised, and these were compared with the figures for those indicators whose schedules remained unchanged. The results suggest that the increase in the maximum performance thresholds improved GPs' performance by 1.77% on average. Low-performing GPs improved significantly more (13.22%) than their high-performing counterparts (0.24%). Changes in maximum performance thresholds are differentially effective in incentivising GPs and could be used further to raise GPs' performance across all indicators.

Contents

1. Introduction

General Practitioners (GPs) in the UK NHS are paid a mixture of capitation, lump sum allowances, items of services and quality incentives under a national contract introduced in 2004 (Morris et al., 2011). The pay-for-performance element, the Quality and Outcomes Framework (QOF), rewards GPs according to their performance against a large number of indicators (ISD Scotland, 2011a). QOF payments represent one of the main sources of potential income for GPs. Participation in the QOF is voluntary in principle (Scottish Government, 2008), but participation is close to universal.

Previous literature suggests that the achievement of quality standards was high when QOF was introduced in 2004 and has increased further since (Steel and Willems, 2010; Vaghela et al., 2009; Doran et al., 2008; Doran et al., 2006). However, it has been argued that the introduction of the QOF payments may have contributed little or nothing to the improvement of quality standards in primary care in the NHS (Doran et al., 2011; Millett et al., 2009; Campbell et al., 2007; Hippisley-Cox et al., 2007). Furthermore, most studies found little association between health outcomes and QOF achievement (Strong et al., 2009; Bottle et al., 2008; Shohet et al., 2007; Williams and de Lusignan, 2006; Fleetcroft and Cookson, 2006).

Under the QOF, practices are paid according to a linear schedule between a lower and upper threshold, which varies across indicators. In April 2006, the indicator set was revised, including the introduction of new indicators, retirement of some indicators and revisions to the definitions of other indicators. Minimum thresholds were raised from 25% to 40% for all indicators. The maximum thresholds were raised for nine clinical indicators while they were left unchanged for 25 other indicators whose definitions remained consistent. This change provides a unique opportunity in the form of a quasi-experiment to evaluate how GPs respond to the changes in threshold payment by comparing the changes of their performance on indicators with increased threshold against those with threshold kept the same.

The changes in the payment threshold under the QOF scheme may have differential effects on different groups of GPs. Therefore, it is important to look beyond the average impact of the policy. We examine whether practices that had initially high and initially low performance responded to the change in the payment threshold differently by dividing them into three subgroups: “Under Performers”, “Competent Performers” and “Excellent Performers”.

The financial incentive for the Under Performers is expected to be the strongest. If they maintain their 2005/6 level of production in 2006/7, the amount of potential income foregone under the QOF scheme is even more than that in 2005/6. Also, the Under Performers have the most room to improve in 2006/7 as they submitted the lowest performance indices among the three groups of GPs. Excellent Performers are expected to respond least to the incentive. There is no direct financial incentive for Excellent Performers to improve their QOF performance. They could achieve the full QOF payment by keeping the same level of performance in 2006/7. They also have the smallest room to improve the QOF performance as they have the highest performance indices in 2006/7. The Competent Performers are expected to respond to the financial incentive positively. The effect may not as strong as on the Under Performers, but should be stronger than on the Excellent Performers. The two empirical tests provide evidence for the above hypotheses.

In summary, the principal aim of this paper is to evaluate the effect of an increase in the maximum performance threshold in the QOF scheme on the performance of GPs in Scotland. We also aim to

examine whether GPs responded to the incentives differently if their performance on the QOF service was different before the introduction of the payment system change.

2. Payment Mechanism

The QOF measures achievement against a range of evidence-based indicators, with points and payments awarded according to the level of achievement. There were 146 QOF indicators in 2005/6. In 2006/7, this was revised to 135 indicators (ISD Scotland, 2011b). This paper focuses on GPs' performance in the clinical domain indicators because most of them have a maximum performance threshold, which will be used to differentiate the high and low performing GPs.

The QOF payment for the clinical domain is the sum of the payments achieved from all the clinical indicators. The calculation of the QOF income from one clinical indicator is explained by equations (1), (2) and (3). The payment is calculated in the same way for all indicators: achieved points multiplied by the value of pounds-per-point.

$$QOF_{igt} = 0 \quad Where \quad 0 \le \alpha_i \le \alpha_l \qquad (1)$$

$$QOF_{igt} = \frac{v_t ADPF_{igt} CPI_{gt} \pi_i^{\max} \times \frac{\alpha_i - \alpha_l}{\alpha_u - \alpha_l}}{\alpha_{iu} D_{igt}} \quad Where\ \alpha_l < \alpha_i < \alpha_u \qquad (2)$$

$$QOF_{igt} = \frac{v_t ADPF_{igt} CPI_{gt} \pi_i^{\max}}{\alpha_{iu} D_{igt}} \quad Where \quad 1 \ge \alpha_i \ge \alpha_u \qquad (3)$$

First, the achieved points are determined by the practice's performance as measured by the performance index (α_i). No points are awarded if the performance index is no more than the minimum performance threshold (α_l) as shown in equation (1). Maximum achievable points ($\pi^{\max}$) are given if the performance index reaches or is above the maximum performance threshold (α_u) as shown in equation (3). Achieved points are linearly correlated to the performance index between the maximum and minimum performance thresholds as shown in equation (2). The performance index for indicator i at year t in practice g is calculated by using the number of patients treated (d_{igt}) out of the number of eligible patients (D_{igt}) as shown in equation (4). Overall the number of points achieved is a piece-wise linear function of the proportion of eligible patients treated as there are two cut-off points at the maximum and minimum performance thresholds.

$$\alpha_{igt} = \frac{d_{igt}}{D_{igt}} \qquad (4)$$

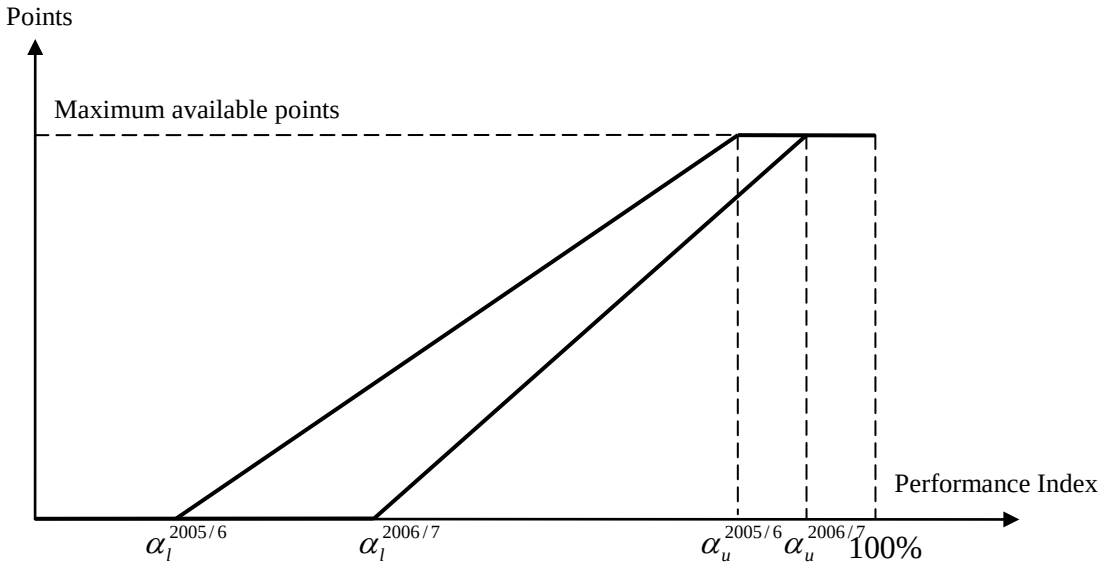
Second, the raw value of pounds-per-point (v) [1] is adjusted by the Contractor Population Index (CPI) [2]. This adjustment gives practices with large lists a high pounds-per-point value. The value also is adjusted by the Adjusted Disease Prevalence Factor ($ADPF$) before it is used to calculate payments. ADPF is used to adjust for the variance in prevalence of disease between practices

[1] The raw value of pounds-per-point was 124.64 pounds in financial years 2005/6 and 2006/7.
[2] CPI is calculated by the list size of the practice against the national average size (set at 5,095 patients for 2004/05 to 2007/08) (ISD Scotland, 2011a).

(Guthrie et al., 2006). A high ADPF value suggests high prevalence of a disease in a practice. After these two adjustments have been made, each general practice will have a unique pounds-per-point value for every disease.

The implication of the policy, which increased the maximum performance thresholds for some indicators in financial year 2006/7, was that GPs were being asked to do more to maintain the same level of QOF income for these altered indicators. The relation between the performance index and achieved points for one QOF indicator, which has the maximum performance threshold increased in 2006/7, is shown in Figure 1.

Figure 1: Performance index and achieved points for one clinical indicator in QOF



3. Method

3.1 Data

The data for the analysis were provided by the Information Services Division (ISD) of the Scottish Government Health Directorates. It includes information that was required for the empirical analysis, i.e. ADPF, maximum achievable points, number of GPs in the practice, the practice population size, the number of eligible patients at indicator level, and the practice performance index. This was for financial years 2005/6 and 2006/7. In Scotland, the majority (approximately 86%) of general practices work under the General Medical Service (GMS) contract (ISD Scotland, 2010). Apart from the GMS contract, general practices can sign up to a Personal Medical Service contract or to the Health Board Primary Medical Services contract. Practices that have opted for non-GMS contracts may opt to use whole or part of QOF as a services quality management tool. The measurements may be tailored in accordance with local circumstances, subject to local agreement. This means the way QOF data is collected in the practices of the non-GMS contract holders is different than how it is collected by GMS contract holders. For this reason, only observations collected from GMS contract practices are used in the empirical analysis. In total there are 40,704

observations from the two financial years. 24 indicators[3] are eligible for this empirical work. The data used in the empirical tests is balanced panel data at practice and indicator.

Practice-indicator observations of indicators where maximum performance thresholds were raised in 2006 can be divided into three groups according to their performance in 2005/6: Under Performers, Competent Performers and Excellent Performers. The performance of the Under Performers in 2005/6 was below the maximum performance threshold for 2005/6. Performance of the Competent Performers was between the maximum performance thresholds of 2005/6 and 2006/7 in 2005/6. Excellent Performers achieved performance indices in 2005/6 that were already higher than the new maximum performance threshold for 2006/7.

3.2 Empirical Framework

The difference-in-differences method is used to evaluate the impact of the policy. This method utilises a quasi-natural experiment. One group of indicators, the treatment group, was subject to a policy change. A second group, with similar characteristics, was not subject to these changes. The second group is assumed to mimic an untreated state and acts as a control in the experiment (Heckman et al., 1999).

A major challenge with the difference-in-differences method is to find an appropriate control group. It is impossible to observe treated subjects in both their treated and untreated state. Therefore, it is necessary to find a proxy for the counterfactual -- that is, how the treated subjects would have performed in their untreated state -- using empirical techniques (Heckman et al., 1999).

We have GPs' performance data for the revised QOF indicators where maximum performance threshold increased in 2006/7 both before and after the change. Two strategies are used to find the proxy for the untreated state.

In the first empirical test, QOF scores for the 17 unrevised clinical indicators are used to proxy the counterfactuals for the three revised clinical indicators. The revised and unrevised indicators are matched at practice level. This method controls for factors that affect the change in performance over time equally across the two groups of indicators. There are 33,896 qualified observations in the

---

[3] Among the 34 indicators (nine clinical indicators have the maximum thresholds increase while 25 clinical indicators whose definitions remain consistent were left unchanged, as explained in the introduction), four of them (CHD8, MH4, MH5 and BP5) have the maximum available points increased in 2006/7. It is difficult to distinguish the effects of two interventions by GPs' performance before and after the policy change. Therefore, the four indicators were excluded from the analysis.

The definitions of five indicators (CHD2, CHD11, CHD12, DM12 and DM15) have no indication about the period of time that the indicators covered. Our analysis required all indicators to cover the same period, 15 months for the first empirical test and nine months for the second empirical test, as the length may positively correlate with the performance indices. Therefore, the five indicators were excluded from the analysis.

BP4 also was excluded from the analysis. Although it has a nine-month treatment period, the service is about record for blood pressure rather than influenza immunisation. Therefore, it is not eligible for the second empirical test.

In total there are 24 indicators eligible for the analysis. 20 of them were analysed in empirical test one. The other four indicators were examined in empirical test two.

first empirical test. They were drawn from 854 practices. The number of observations in the control group is 28,960 over the two years.

In the second empirical test, the strategy is to match each revised clinical indicator with one or more unrevised indicator(s) by practice and type of clinical service. This means the controls are restricted to observations which were identical to the treatment group in terms of practice and form of treatment provided, in this case, influenza immunisation. Four indicators were selected for the second empirical test: CHD12, COPD8, DM18 and STROKE10. This method reduces the differences between the treatment and control group even further. In total there are 6,808 observations were drawn from 853 practices covering the two financial years. The number of observations in the control group is 5,102.

Three methods are used to control for constant, practice level characteristics that may confound our findings. First, we match observations of unrevised QOF indicators with observations of revised QOF indicators by practice. In other words, GPs' performance for the unrevised QOF indicators in 2006/7 is assumed to mimic the untreated state.

Second, we include a number of exogenous variables (i.e. a group of financial factors) to adjust for time variant differences between the treatment and control group.

Thirdly, we employ a fixed effects method to control any other unobservable time-invariant practice characteristics.

3.3 Empirical test one: comparable indicators

3.3.1 Econometric model

The first empirical test uses 20 QOF indicators. The definitions are presented in Table 1. DM7, DM17, and ASTHMA3 had their maximum performance threshold increased in 2006. The remaining 17 indicators were unchanged.

Table 1: Description of the 20 indicators in 2005/6 and 2006/7

Indicator names	Maximum threshold 2005/6	Maximum threshold 2006/7	Maximum achievable points	Definitions
DM7	**85%**	**90%**	**11**	**The percentage of patients with diabetes in whom the last HbA1c is 10 or less (or equivalent test/reference range depending on local laboratory) in the previous 15 months.**
DM17	**60%**	**70%**	**6**	**The percentage of patients with diabetes whose last measured total cholesterol within previous 15 months is 5 mmol/l or less.**
ASTHMA3	**70%**	**80%**	**6**	**The percentage of patients with asthma between the ages of 14 and 19 in whom there is a record of smoking status in the previous 15 months.**

Indicator names	Maximum threshold 2005/6	Maximum threshold 2006/7	Maximum achievable points	Definitions
ASTHMA6	70%	70%	20	The percentage of patients with asthma who have had an asthma review in the previous 15 months.
CHD5	90%	90%	7	The percentage of patients with coronary heart disease whose notes have a record of blood pressure in the previous 15 months.
CHD6	70%	70%	19	The percentage of patients with coronary heart disease in whom the last blood pressure reading (measured in the previous 15 months) is 150/90 or less.
CHD7	70%	70%	7	The percentage of patients with coronary heart disease whose notes have a record of total cholesterol in the previous 15 months.
CHD9	70%	70%	7	The percentage of patients with coronary heart disease with a record in the previous 15 months that aspirin, an alternative anti-platelet therapy, or an anti-coagulant is being taken (unless a contraindication or side-effects are recorded).
DM2	90%	90%	3	The percentage of patients with diabetes whose notes record BMI in the previous 15 months.
DM5	90%	90%	3	The percentage of diabetic patients who have a record of HbA1c or equivalent in the previous 15 months.
DM9	90%	90%	3	The percentage of patients with diabetes with a record of the presence or absence of peripheral pulses in the previous 15 months.
DM10	90%	90%	3	The percentage of patients with diabetes with a record of neuropathy testing in the previous 15 months.
DM11	90%	90%	3	The percentage of patients with diabetes who have a record of the blood pressure in the previous 15 months.
DM13	90%	90%	3	The percentage of patients with diabetes who have a record of micro-albuminuria testing in the previous 15 months (exception reporting for patients with proteinuria).
DM16	90%	90%	3	The percentage of patients with diabetes who have a record of total cholesterol in the previous 15 months.
STROKE5	90%	90%	2	The percentage of patients with TIA or stroke who have a record of blood pressure in the notes in the preceding 15 months.
STROKE6	70%	70%	5	The percentage of patients with a history of TIA or stroke in whom the last blood pressure reading (measured in the previous 15 months) is 150/2 or less.
STROKE7	90%	90%	2	The percentage of patients with TIA or stroke who have a record of total cholesterol in the

Indicator names	Maximum threshold 2005/6	Maximum threshold 2006/7	Maximum achievable points	Definitions
				last 15 months.
STROKE8	60%	60%	5	The percentage of patients with TIA or stroke whose last measured total cholesterol (measured in the previous 15 months) is 5 mmol/l or less.
THYROI2	90%	90%	6	The percentage of patients with hypothyroidism with thyroid function tests recorded in the previous 15 months.

We evaluate the effect of the policy change by comparing GPs' performance in the 17 indicators with their performance in the other three indicators before and after the change was implemented.

The model may be specified as follows:

$$y_{igt} = \beta_1 + (X_{igtI} X_t)\beta_2 + X_{igtI}\beta_3 + X_t\beta_4 + X_{igtv}\beta_5 + \varepsilon_{igt} \qquad (5)$$

The dependent variable is GPs' performance index (y_{igt}) [4], measured by the number of patients treated for indicator i at year t in practice g out of the number of eligible patients.

Treatment (X_{igtI}) is a dummy variable to indicate whether an observation was drawn from the treatment group or the control group. Indicators DM7, DM17 and ASTHMA3 are defined as the treatment group. Observations from the other 17 indicators are categorised in the control group with a value of zero for this variable.

Year dummy (X_t) is one if the observations come from 2006/7. Otherwise, it is zero.

The difference-in-differences variable $(X_{igtI} X_t)$ is defined as the combined effect of the year dummy and the treatment group. It is created by multiplying the treatment group and the year dummy. The coefficient (β_2) reflects the policy effect. The basic form of the difference-in-differences estimator of the average treatment effect (ATE) is explained in equation (6). It compares mean outcomes for the treatment group and control group before and after the treatment.

$$ATE_{DiD} = (\bar{y}^{Treated}_{After} - \bar{y}^{Treated}_{Before}) - (\bar{y}^{Control}_{After} - \bar{y}^{Control}_{Before})$$

$$= [(\beta_1 + \beta_2 + \beta_3 + \beta_4 + X^{Treated}_{igv(after)}\beta_5) - (\beta_1 + \beta_3 + X^{Treated}_{igv(before)}\beta_5)] - [(\beta_1 + \beta_4 + X^{Control}_{igv(after)}\beta_5) - (\beta_1 + X^{Control}_{igv(before)}\beta_5]$$

$$= \underbrace{\beta_2}_{\text{Policy Effect}} + \underbrace{[(X^{Treated}_{igv(after)}\beta_5) - (X^{Treated}_{igv(before)}\beta_5)] - [(X^{Control}_{igv(after)}\beta_5) - (X^{Control}_{igv(before)}\beta_5]}_{\text{The effect is captured by } X_{igtv}} \qquad (6)$$

[4] It equals to α_{igt} in equation (4).

X_{igtv} is a $n \times 4$ matrix for observable time variant variables. The variables included in this matrix are thousands of patients per GP, number of patients that need to be treated to get the full payment $(\alpha_u D_{igt})$, ADPF and maximum achievable points $(\pi^{\max})$ [5].

ε_{git} is the error term, which includes all the other unobserved variables and random errors.

The fixed effect method is applied to control for the unobserved time invariant practice characteristics.

3.3.2 Regressions on the three subsamples

The empirical test first estimates the model using all observations and then splits the sample into three groups: Under Performers, Competent Performers and Excellent Performers. In the three subsample regressions, there are 158,744 and 4,034 observations for the three treatment groups respectively. All the observations under the 17 untreated indicators are included for the control group, regardless of indicator-practice performance in 2005/6. The summary statistics are presented in Table 2.

[5] $\pi^{\max}$ could suggest the level of difficulty to achieve the QOF rewards. An indicator may have more points because it is more difficult to achieve. A negative coefficient is therefore expected. However, $\pi^{\max}$ could also suggest the potential level of financial rewards for GPs as it is positively associated with the potential QOF income. In that case, a positive coefficient is expected.

Table 2: Summary of statistics for the 20 selected indicators for the three groups of performers in years 2005/6 and 2006/7[1]

Groups of performers			Under Performers				Competent Performers				Excellent Performers	
Variables	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Treatment group for financial year 2005/6												
Performance index	0.72	0.12	0.41	0.85	0.79	0.09	0.61	0.90	0.89	0.07	0.70	1
$\alpha_u D_{igt}$	72.11	69.08	5	287	93.40	89.80	5	483	92.60	81.49	4	513
ADPF	1.03	0.13	0.79	1.48	1.01	0.12	0.47	1.61	1.00	0.11	0.47	1.53
Maximum achievable points	8.85	2.49	6	11	7.99	2.45	6	11	7.64	2.35	6	11
Thousands of patients per GP	1.28	0.53	0.13	2.60	1.32	0.45	0.27	3.37	1.31	0.48	0.13	3.80
Observations				79				372				2017
Control group for financial year 2005/6												
Performance index	0.88	0.11	0.40	1	0.91	0.09	0.41	1	0.93	0.08	0.40	1
$\alpha_u D_{igt}$	117.49	92.09	3	580	172.26	121.08	5	830	151.54	109.24	3	830
ADPF	0.99	0.14	0.45	1.77	1	0.13	0.39	1.65	1	0.14	0.39	1.77
Maximum achievable points	5.94	5.22	2	20	5.94	5.22	2	20	5.94	5.23	2	20
Thousands of patients per GP	1.28	0.53	0.13	2.60	1.33	0.44	0.27	3.37	1.30	0.49	0.13	3.80
Observations				1203				5050				14111
Treatment group for financial year 2006/7												
Performance index	0.88	0.08	0.61	1	0.88	0.08	0.50	1	0.91	0.06	0.53	1
$\alpha_u D_{igt}$	76.96	73.41	5	304	103.59	99.23	5	540	106.50	91.96	4	538
ADPF	1.03	0.11	0.80	1.32	1.01	0.11	0.40	1.58	1.01	0.11	0.40	1.57
Maximum achievable points	8.85	2.49	6	11	7.99	2.45	6	11	7.64	2.35	6	11
Thousands of patients per GP	1.28	0.52	0.13	2.56	1.32	0.45	0.26	3.41	1.31	0.48	0.13	3.72
Observations				79				372				2017
Control group for financial year 2006/7												
Performance index	0.91	0.10	0.20	1	0.93	0.08	0.44	1	0.94	0.07	0.20	1
$\alpha_u D_{igt}$	118.35	89.47	3	563	175.60	121.96	6	851	155.43	110.63	3	851
ADPF	1	0.13	0.42	1.74	1	0.13	0.36	1.58	1.01	0.13	0.36	1.74
Maximum achievable points	5.94	5.22	2	20	5.94	5.22	2	20	5.94	5.23	2	20
Thousands of patients per GP	1.28	0.52	0.13	2.56	1.33	0.44	0.26	3.41	1.31	0.49	0.13	3.72
Observations				1203				5050				14111

[1] Eligible practices should have at least one control indicator and one treatment indicator. Therefore, some practices have less than 20 indicators in our dataset. The means are over practices and indicators.

It should be noted that there is a potential problem with the control group subsamples. The selection of observations in the control group does not take into account whether the performance index falls above or below the maximum performance threshold. GPs were incentivised in 2006/7 to switch their efforts from indicators in the control group to those where it had not. For GPs who have already achieved the upper performance thresholds in all the control group indicators, this effect should be most acute. However, if the control group is dominated by GPs who had not achieved the upper threshold in 2005/6, then the choice between whether they would choose to raise their performance in the control group indicators or the treatment group indicators is unclear. If they had chosen to spend their effort in catching up with the performance in the control group indicators rather than on keeping pace with the increased demands of the treatment indicators, then the estimated effect of the policy may well be negative. The problem may be specified as follows:

$$ATE_{DiD} = \beta_2 = (\bar{y}^{Treated}_{After} - \bar{y}^{Treated}_{Before}) - (\bar{y}^{Control}_{After} - \bar{y}^{Control}_{Before}) < 0 \qquad (7)$$

$$\text{when} \qquad \bar{y}^{Treated}_{After} - \bar{y}^{Treated}_{Before} \geq 0$$

For this reason, a further regression was implemented using the Excellent Performers subsample. This time, the observations in the control group were limited to those who had achieved the maximum performance threshold in 2005/6. The problem outlined above was thereby avoided. The statistical summary for this dataset is presented in Table 3.

Table 3: Summary of statistics for the 20 selected indicators in years 2005/6 and 2006/7 with Excellent Performers in both control group and treatment group[1]

Variables	Mean	SD	Min	Max
Treatment group for financial year 2005/6				
Performance index	0.89	0.07	0.70	1
$\alpha_u D_{igt}$	92.60	81.49	4	513
ADPF	1	0.11	0.47	1.53
Maximum achievable points	7.64	2.35	6	11
Thousands of patients per GP	1.31	0.48	0.13	3.80
Observations				2017
Control group for financial year 2005/6				
Performance index	0.94	0.07	0.60	1
$\alpha_u D_{igt}$	150.88	109.61	3	830
ADPF	1	0.14	0.39	1.77
Maximum achievable points	6.11	5.32	2	20
Thousands of patients per GP	1.30	0.49	0.13	3.80
Observations				12831
Treatment group for financial year 2006/7				
Performance index	0.09	0.06	0.53	1
$\alpha_u D_{igt}$	106.50	91.96	4	538
ADPF	1.01	0.11	0.40	1.57
Maximum achievable points	7.64	2.35	6	11
Thousands of patients per GP	1.31	0.48	0.13	3.72
Observations				2017
Control group for financial year 2006/7				
Performance index	0.94	0.06	0.38	1

$\alpha_u D_{igt}$	154.89	111.03	3	851
ADPF	1.01	0.13	0.36	1.74
Maximum achievable points	6.11	5.32	2	20
Thousands of patients per GP	1.30	0.49	0.13	3.72
Observations				12831

[1]Eligible practices should have at least one control indicator and one treatment indicator. Therefore, some practices have less than 20 indicators in our dataset. The means are over practices and indicators.

3.4 Empirical test two: matched indicators

Four indicators are included in this empirical test. Their definitions are reported in Table 4. These four indicators aim to improve the rate of influenza vaccination. The only indicator that had its maximum performance threshold increased in 2006 was CHD12; COPD8, DM18 and STROKE10 were unchanged. The specification for this model is the same as equation (5), and the difference-in-differences estimator with fixed effects method is again employed.

Table 4: Definitions of four indicators relating to GPs' record keeping for influenza immunisation

Names	Maximum threshold 2005/6	Maximum threshold 2006/7	Maximum achievable points	Definitions
CHD12	**85%**	**90%**	**7**	**The percentage of patients with coronary heart disease who have a record of influenza immunisation in the preceding 1st September to 31st March.**
COPD8	85%	85%	6	The percentage of patients with COPD who have had influenza immunisation in the preceding 1st September to 31st March.
DM18	85%	85%	3	The percentage of patients with diabetes who have had influenza immunisation in the preceding 1st September to 31st March.
STROKE10	85%	85%	2	The percentage of patients with TIA or stroke who have had influenza immunisation in the preceding 1st September to 31st March.

As before, the practices are split into three groups: Under Performers, Competent Performers and Excellent Performers. The empirical test first estimates the model using all observations, and then splits the sample into three groups as described above. In the three subsample regressions have 90, 276 and 1,340 observations, respectively, for the three treatment groups. The summary statistics are presented in Table 5.

Table 5: Summary of statistics for the four selected indicators for the three groups of performers in years 2005/6 and 2006/7[1]

Groups of performers			Under Performers				Competent Performers				Excellent Performers	
Variables	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Treatment group for financial year 2005/6												
Performance index	0.79	0.06	0.61	0.85	0.88	0.01	0.85	0.90	0.96	0.03	0.90	1
$\alpha_u D_{igt}$	151.09	105.12	18	418	233.58	125.18	34	713	173.16	109.76	5	707
ADPF	1.03	0.14	0.70	1.43	1.01	0.12	0.41	1.32	1	0.14	0.41	1.59
Maximum achievable points	7	0	7	7	7	0	7	7	7	0	7	7
Thousands of patients per GP	1.34	0.56	0.27	2.86	1.30	0.44	0.30	2.74	1.29	0.50	0.13	3.80
Observations				45				138				670
Control group for financial year 2005/6												
Performance index	0.81	0.11	0.50	1	0.89	0.05	0.68	1	0.95	0.05	0.50	1
$\alpha_u D_{igt}$	78.37	64.85	7	329	120.21	76.45	9	492	88.91	64.01	5	419
ADPF	1.03	0.19	0.57	1.69	1.01	0.17	0.39	1.71	1.01	0.19	0.39	2.07
Maximum achievable points	3.68	1.71	2	6	3.67	1.70	2	6	3.66	1.70	2	6
Thousands of patients per GP	1.33	0.55	0.27	2.86	1.30	0.44	0.30	2.74	1.30	0.49	0.13	3.80
Observations				134				414				2003
Treatment group for financial year 2006/7												
Performance index	0.87	0.08	0.63	1	0.91	0.05	0.69	1	0.95	0.04	0.73	1
$\alpha_u D_{igt}$	151.82	107.62	17	433	237.04	126.76	25	738	181.05	113.49	5	697
ADPF	1.05	0.13	0.77	1.40	1.02	0.11	0.64	1.33	1	0.14	0.36	1.52
Maximum achievable points	7	0	7	7	7	0	7	7	7	0	7	7
Thousands of patients per GP	1.34	0.55	0.26	2.88	1.31	0.44	0.30	2.86	1.30	0.50	0.13	3.72
Observations				45				138				670
Control group for financial year 2006/7												
Performance index	0.86	0.10	0.52	1	0.89	0.07	0.55	1	0.94	0.05	0.50	1
$\alpha_u D_{igt}$	77.13	62.74	6	318	120.65	76.39	11	454	90.54	65.34	5	439
ADPF	1.04	0.18	0.64	1.81	1.02	0.16	0.62	1.74	1.01	0.18	0.40	2.12
Maximum achievable points	3.68	1.71	2	6	3.67	1.70	2	6	3.66	1.70	2	6
Thousands of patients per GP	1.33	0.54	0.26	2.88	1.31	0.44	0.30	2.86	1.30	0.49	0.13	3.72
Observations				134				414				2003

[1]Eligible practices should have at least one treatment indicator. Therefore, some practices have less than three treatment indicators in our dataset. The means are over practices and indicators.

An additional regression again was conducted on the Excellent Performers subsample. Observations from the control group were limited to those who had achieved the maximum performance threshold in 2005/6. The summary statistics are presented in Table 6.

Table 6: Summary of statistics for the four selected indicators in years 2005/6 and 2006/7 with Excellent Performers in both control group and treatment group[1]

Variables	Mean	SD	Min	Max
Treatment group for financial year 2005/6				
Performance index	0.96	0.03	0.90	1
$\alpha_u D_{igt}$	173.16	109.76	5	707
ADPF	1	0.14	0.41	1.59
Maximum achievable points	7	0	7	7
Thousands of patients per GP	1.29	0.50	0.13	3.80
Observations				670
Control group for financial year 2005/6				
Performance index	0.95	0.04	0.85	1
$\alpha_u D_{igt}$	89.13	64.17	5	419
ADPF	1.01	0.19	0.39	2.07
Maximum achievable points	3.69	1.70	2	6
Thousands of patients per GP	1.30	0.49	0.13	3.80
Observations				1947
Treatment group for financial year 2006/7				
Performance index	0.95	0.04	0.73	1
$\alpha_u D_{igt}$	181.05	113.49	5	697
ADPF	1	0.14	0.36	1.52
Maximum achievable points	7	0	7	7
Thousands of patients per GP	1.30	0.50	0.13	3.72
Observations				670
Control group for financial year 2006/7				
Performance index	0.94	0.05	0.65	1
$\alpha_u D_{igt}$	90.88	65.57	5	439
ADPF	1.01	0.18	0.40	2.12
Maximum achievable points	3.69	1.70	2	6
Thousands of patients per GP	1.30	0.49	0.13	3.72
Observations				1947

[1] Eligible practices should have at least one treatment indicator. Therefore, some practices have less than three treatment indicators in our dataset. The means are over practices and indicators.

4. Results

4.1 Empirical test one: comparable indicators

The results are presented in column two in Table 7. The interaction effect is statistically significant[6] and positive in the first empirical test. It suggests that the increase in the maximum performance threshold increased GPs' performance on the three intervention indicators by 1.77%. This is the difference-in-differences effect. The year dummy is statistically significant and positive. The result shows that GPs' overall performance was 0.86% higher in 2006/7 than in 2005/6. The performance of the control group is 2.38% higher than that of the treatment group, which is statistically significant. This coefficient captures the average difference in the level performance of the two groups excluding the effect of the policy change. For every extra patient needed to be treated to achieve the full QOF points, overall performance increases by 0.04%. ADPF is negatively correlated with GPs' overall performance. The maximum achievable points is negatively correlated with GPs' performance. Each extra point on average decreases GPs' overall performance of QOF services by 0.49%. Increasing the number of patients per GP reduces GPs' overall performance by 3.81%. GPs with more patients registered in their practice had poorer performance.

[6] This is at the 5% level as are all the statistically significant coefficients in this paper.

Table 7: Determinants of the performance index in 2005/6 and 2006/7 modelled for the 20 selected indicators – overall sample and three subsamples from empirical test one

Independent variables	Overall	t	Under Performers	t	Competent Performers	t	Excellent Performers	t	Excellent Performers vs Excellent Performers	t
Interaction effect	0.0177	8.41	0.1322	8.48	0.0688	12.06	0.0024	1.09	0.0078	4.12
Year dummy	0.0086	10.74	0.0337	8.71	0.0165	11.01	0.0074	9.47	0.0018	2.51
Treatment dummy	-0.0238	-15.38	-0.1145	-10.09	-0.0759	-18.16	-0.0093	-5.73	-0.0186	-13.46
$\alpha_u D_{igt}$	0.0004	58.77	0.0006	14.80	0.0004	33.39	0.0004	58.97	0.0004	70.18
ADPF	-0.0906	-20.02	-0.1298	-6.77	-0.0634	-7.42	-0.0928	-20.60	-0.0905	-22.44
Maximum achievable points	-0.0049	-62.82	-0.0053	-13.95	-0.0055	-37.51	-0.0050	-65.87	-0.0054	-80.93
Thousands of patients per GP	-0.0381	-3.28	-0.3797	-5.70	-0.0680	-2.41	-0.0294	-2.58	-0.0213	-2.14
Constant	1.0400	65.49	1.4576	16.54	1.0323	26.7	1.0336	66.27	1.0343	75.81
Observations for treatment group		4936		158		744		4034	4034	
Observations for control group		28960		2406		10100		28222	12831	

For the three subsamples, all the observations under the 17 untreated indicators are included for the control group, regardless of indicator-practice performance in 2005/6.

4.2 Empirical test one: subsample results

Column four and column six of Table 7 report the results from the Under Performers and Competent Performers respectively. Results from the fourth column show the effect of the increased maximum performance threshold on GPs who did not reach the maximum performance threshold in 2005/6. The effect of the change in threshold is statistically significant and positive to increase the underperforming GPs' performance on the three intervention indicators by 13.22%. This is the difference-in-differences effect. The results from column six show the effect of the increased maximum performance threshold on GPs whose performance index was located between the maximum performance thresholds of 2005/6 and 2006/7 in 2005/6. The effect of the policy is to increase the competent GPs' performance on the three intervention indicators by 6.88%.

Results from the eighth column show that the interaction effect is not statistically significant for Excellent Performers in the first empirical test. It suggests that the policy of increasing the maximum performance threshold for the three clinical indicators had no significant effect on the Excellent Performers.

The year dummy is statistically significant and positive. GPs' overall performance was 0.74% higher in 2006/7 than in 2005/6. Their performance is 0.93% lower in the treatment group than in the control group; this is statistically significant. Raising the upper threshold significantly increases the number of patients need to be treated to achieve the full QOF points $(\alpha_u D_{igt})$. For every extra patient GPs must treat in order to achieve the full QOF points, the overall performance increases by 0.04%, identical to the results derived from empirical test one by using the pooled data. ADPF and the number of points available are both negatively correlated with GPs' overall performance. Each extra point on average decreases GPs' overall performance under QOF by 0.5%. The effect of raising patient numbers per GP is to decrease GPs' overall performance by 2.94%.

Column ten of Table 7 reports the effect of the increased payment threshold on the Excellent Performers. The difference between this estimation and the estimation used to produce the result reported in column eight is that here, only observations with a performance index above the 2005/6 maximum performance threshold are included in the control group. The results reported in Table 7 suggest that the effect of the policy is to significantly increase GPs' performance on the three intervention indicators by 0.78%.

4.3 Empirical test two: matched indictors

The results are presented in the second column of Table 8. The interaction effect is positive in the second empirical test. The results suggest that the increase in maximum performance threshold increased GPs' performance on the intervention indicator by 0.86%. This is the difference-in-differences effect. The year dummy is statistically significant and negative. The results show that GPs' overall performance was 0.34% lower in 2006/7 than it was in 2005/6. The performance of the control group is 1.44% higher than that of the treatment group, which is statistically significant. ADPF is negatively correlated with GPs' overall performance. The number of maximum achievable points is positively correlated with GPs' performance. Each extra point on average increases GPs' overall performance under QOF by 0.68%.

Table 8: Determinants of the performance index in 2005/6 and 2006/7 modelled for the four selected indicators – overall sample and the three subsamples from empirical test two

Independent variables	Overall	t	Under Performers	t	Competent Performers	t	Excellent Performers	t	Excellent Performers vs Excellent Performers	t
Interaction effect	0.0086	3.68	0.0417	2.10	0.0260	4.35	0.0027	1.21	0.0046	2.32
Year dummy	-0.0034	-2.93	0.0452	4.52	0.0049	1.60	-0.0084	-7.42	-0.0103	-10.16
Treatment dummy	-0.0144	-5.94	-0.0802	-3.94	-0.0414	-6.47	-0.0053	-2.28	-0.0067	-3.18
$\alpha_u D_{igt}$	-1.35e-5	-1.06	0.0001	1.10	-1.18e-5	-0.42	-1.25e-5	-0.98	-9.19e-6	-0.80
ADPF	-0.0231	-4.84	-0.0228	-0.59	-0.0233	-1.88	-0.0277	-5.97	-0.0324	-7.72
Maximum achievable points	0.0068	19.30	0.0144	4.59	0.0098	10.90	0.0056	16.68	0.0049	16.02
Thousands of patients per GP	-0.0018	-0.12	-0.0593	-0.24	-0.0425	-0.65	0.0073	0.51	0.0075	0.59
Constant	0.9316	43.92	0.8497	2.58	0.9344	10.84	0.9444	49.70	0.9553	56.30
Observations for treatment group		1706		90		276		1340	1340	
Observations for control group		5102		268		828		4006	3894	

For the three subsamples, all the observations under the three untreated indicators are included for the control group, regardless of indicator-practice performance in 2005/6.

4.4 Empirical test two: subsample results

The results from the fourth and sixth columns of Table 8 refer to Under Performers and Competent Performers, respectively. In both estimations, the coefficients of interaction effect are significantly positive, suggesting that the 2006/7 increase in maximum performance threshold raised the QOF performance of these two groups. The performance of the Under Performers on the intervention indicator increased by 4.17%, while that of the Competent Performers improved by 2.6%. This is the difference-in-differences effect.

Results from the eighth column show that the interaction effect is not statistically significant for Excellent Performers in this empirical test. It suggests that the policy of increasing the maximum performance threshold for clinical indicator CHD12 had no significant effect on the Excellent Performers. The result shows that GPs' overall performance was 0.84% lower in 2006/7 than it was in 2005/6. GPs' performance under QOF is 0.53% lower in the treatment group than in the control group; the difference is statistically significant. ADPF is statistically significant and negatively correlated with GPs' overall performance. The number of maximum achievable points is positively correlated with GPs' overall performance. Each extra point on average increases GPs' overall performance under QOF by 0.56%.

Column ten of Table 8 reports the effect of the increased payment threshold on the Excellent Performers. The difference between this estimation and the estimation used to produce the result reported in column eight is that only observations with a performance index above the 2005/6maximum performance threshold are included in the control group. The results reported in Table 8 suggest that the raising of the threshold significantly increased GPs' performance on the intervention indicator by 0.46%.

5. Discussion

The coefficients of the interaction effect terms established in empirical tests one and two suggest that the increase in the maximum performance threshold had a positive effect on GPs' QOF performance. These results provide empirical evidence that the 2006 change in QOF policy did motivate GPs. The pooled results have been confirmed by the results derived from the subsamples in the two empirical tests. We also find that the Under Performers improved the most in 2006/7. The main findings are summarised in Table 9.

Table 9: The coefficients of the interaction effect (the effect of the 2006 policy change on GPs' performance) in the two empirical tests

	Overall observations	Under Performers	Competent Performers	Excellent Performers	Excellent Performers vs Excellent Performers
Empirical test one	1.77%	13.22%	6.88%	0.24%	0.78%
Empirical test two	0.86%	4.17%	2.60%	0.27%	0.46%

The coefficients for the Excellent Performers in the fifth column are not statistically significant. The rest of the coefficients in Table 9 are statistically significant at the 5% level.

These findings are supported by Kontopantelis et al. (2011). They investigate the effect on GP influenza immunization rates of setting higher targets in the QOF scheme in England by comparing the changes in GPs' achievement for patients with CHD against those with COPD, Diabetes and stroke. Our second empirical test evaluates the same effect. We examine the same four chronic diseases (CHD, COPD, diabetes and stroke) and treatment (influenza immunisation) in empirical test two.

These two papers differ in study setting, method and data. Our paper evaluates the increased maximum performance threshold of the QOF scheme in Scotland on the performance of GPs in the influenza immunisation service. We also evaluate the impact of the same policy on GPs' performance across a wider range of treatments. Kontopantelis et al. (2011) examines the effect of setting higher QOF targets on GPs' achievement rates and exception reporting rates[7], specifically for influenza immunisation. Furthermore, there are two main differences between the two studies in method. First, Kontopantelis et al. (2011) employs a longitudinal random-effect multi-level linear regression, while this paper uses a difference-in-differences estimator with a fixed effects method. Second, this paper has more independent variables than Kontopantelis et al. (2011). It includes a number of financial factors as well as the different-in-differences variable. However, those factors are not controlled for in Kontopantelis et al. (2011). In addition to the differences in setting and method, the data used in the two studies are different. This paper examines the impact of the QOF in Scotland using 2005/6 and 2006/7 data. Kontopantelis et al. (2011) examine the impact in England using a longer series of QOF data from 2004/5 to 2009/10.

Despite these differences in setting, method and data, the two studies find the same general effect, i.e. that compared to their performance for the other three indicators, practice performance improved in 2006/7 for indicator CHD12. The effect was observed across Under Performers, Competent Performers and Excellent Performers. Furthermore, both studies find that the increase was largest for Under Performers and smallest for Excellent Performers. We also find that raising the threshold increased Excellent Performers' achievement for indicator CHD12 by 0.46%, which is similar with found by Kontopantelis et al. (2011). They find the magnitude of the CHD12 upper threshold change on CHD12 outcome for practices with 2005/6 CHD12 reported achievement[8] above 90% (Excellent Performers) is at 0.49%.

The positive relationship between the increase in the QOF incentive payment and GPs' performance also is supported by other studies. Doran et al. (2008) provide evidence that the QOF scheme made a useful contribution to improving GPs' performance. They find that the median overall reported QOF performance index was increased from 85.1% in 2004/5 to 89.3% in 2005/6 and 90.8% in 2006/7. Vaghela et al. (2009) investigate the effect of QOF incentives on diabetes outcomes. Diabetes is the most studied QOF condition (Steel and Willems, 2010). That study analysed 98% of all English general practices between 2004/5 and 2007/8. The result suggests that the overall level of achievement of diabetes targets increased over the four years.

Under Performers

The positive effect of the increased maximum performance threshold on the QOF outputs of Under Performers suggests that the marginal income they derived from QOF payment is greater than the

[7] Exception reporting rate is calculated by the number of patients who are on the relevant QOF register but not included in the calculation of a practice's QOF achievement against the number of patients that are registered for QOF service. The reasons that a patient might be reported as an exception include: the treatment not being clinically appropriate for the patient, the patient not appearing for treatment, the patient refusing to have the treatment, or the patient only having been diagnosed/registered with the practice very recently.

[8] It is equal to the performance index in this paper.

marginal cost in 2005/6. If GPs' performance is at equilibrium in 2005/6, it suggests that the marginal income from QOF work is equal to the marginal costs. The QOF income from treating each patient decreased as a result of the increase in the maximum performance threshold in 2006/7. The total QOF income did not change over time. However, GPs need to treat more patients in 2006/7 to keep the same amount of QOF payment. The result also suggests that the decreased marginal utility remained higher than marginal cost in 2006/7, incentivising GPs to raise their provision of QOF services in 2006/7.

Excellent Performers

The majority of the observations relate to Excellent Performers (95.16%[9] in the first empirical test and 78.53%[10] in the second empirical test). They are unable to derive more income from the QOF scheme by maintaining their performance level in 2006/7. The achieved performance indices for the Excellent Performer in 2005/6 were already higher than the new maximum performance threshold for 2006/7. Therefore, the policy change had an impact on GPs' non-QOF incentivised activities first in 2006.

The positive effect of the increased maximum performance threshold on Excellent Performers could be explained by two things. Firstly, the increase fostered the marginal utility that GPs derived from the non-QOF incentivised activities. Secondly, the policy was undermined their marginal utility that derived from treating patients beyond the maximum performance threshold. However, the undermining effect was outweighed by the decrease in marginal costs. The marginal costs of supplying the non-QOF incentivised service include material costs, effort costs and forgone income. The material costs did not change. The effort required to produce non-QOF incentivised activities with the increase in output. The marginal forgone income decreased in 2006/7 if we take the marginal income from the same QOF service as a proxy[11]. This paper is unable to test which explanation is most likely.

Competent performers

The effect of this policy change among Competent Performers has been significantly positive, although some GPs may not have responded promptly to the change. GPs who did realise that the maximum performance threshold increased in 2006 may have responded in the same way as the Under Performers. Otherwise, the explanation for this result is the same as discussed above for Excellent Performers.

Limitations

The fixed effects method cannot control for the effect of unobserved time variant patient characteristics. This could bias the results. For example, GPs may provide fewer QOF services to patients with serious health conditions and co-morbidities, but more to relatively healthier patients

[9] The percentage of Excellent Performer in the pooled sample in the first empirical test is 95.16%. The data used is derived from Table 7. It is calculated as

$$\frac{32256}{33896} = 95.16\%$$

[10] The percentage of Excellent Performer in the pooled sample in the second empirical test is 78.53%. The data used is derived from Table 8. It is calculated as

$$\frac{5346}{6808} = 78.53\%$$

[11] The reason of having a decrease in the marginal QOF income as the result of the increased upper threshold of the QOF scheme is explained in the previous section (for the Under Performers).

or patients with relatively straightforward health problems. This limitation is particularly relevant here because all the treatment group indicators in the two experiments are ongoing management clinical indicators (DM7, DM17 and ASTHMA3 in the first empirical test and CHD12 in the second empirical test). These indicators address GPs' provision of clinical treatment rather than their record-keeping skills; thus, the health status of their patients is likely to have a greater impact on their measured performance under these indicators.

Furthermore, this paper does not exclude the possibility that GPs exploited the QOF scheme to increase their income. Following their examination of data gathered from Scottish general practices in 2004/5, Gravelle, Sutton and Ma (2010) suggested that the QOF payment system creates incentives for GPs to manipulate their reporting of exception rates. Evidence of gaming also is found in Kontopantelis et al. (2011). Their result suggests that a 5% increase in the upper payment threshold for CHD was associated with a 0.26% increase in the proportion of CHD patients exceptions reported. This paper does not take into account the effect of gaming.

6. Conclusion

The principal finding of this paper is that the increased maximum performance threshold under the QOF scheme in 2006 was differentially effective in improving GPs' performance. Overall, the effect of the policy change in 2006 improved the performance of GPs in Scotland. We also conclude that the positive effect of the policy change on GPs' performance under QOF is significantly stronger for low-performing GPs than for high performers. These empirical findings imply that raising the QOF indicator threshold further may incentivise GPs' performance across all indicators.

References

Bottle, A., Gnani, S., Saxena, S., Aylin, P., Mainous, A.G. and Majeed, A. (2008) Association between quality of primary care and hospitalization for coronary heart disease in England: National cross-sectional study. *Journal of General Internal Medicine.* 23(2), 135-141.

Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B. and Roland, M. (2007) Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine*. 357(2), 181–190.

Doran, T., Kontopantelis, E., Valderas, J.M., Campbell, S., Roland, M., Salisbury, C. and Reeves, D. (2011) Effect of financial incentives on incentivised and non-incentivised clinical activities: Longitudinal analysis of data from the UK Quality and Outcomes Framework. *British Medical Journal*. 342, d3590.

Doran, T., Fullwood, C., Kontopantelis, E. and Reeves, D. (2008) Effect of financial incentives on inequalities in the delivery of primary clinical care in England: Analysis of clinical activity indicators for the QOF. *The Lancet*. 372(9640), 728-736.

Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U. and Roland, M. (2006) Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine*. 355(4), 375-384.

Fleetcroft, R. and Cookson, R. (2006) Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Service Research & Policy.* 11(1), 27-31.

Guthrie, B., McLean, G. and Sutton, M. (2006) Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. *British Journal of General Practice.* 56(532), 836-841.

Heckman, J.J., Lalonde, R.J. and Smith, J.A. (1999) The economics and econometrics of active labor market programs. *In* Ashenfelter, O. and Layard, R. eds. *Handbook of labor economics.* Vol. 3, Part A. Elsevier: The Netherlands. 1865-2097.

Hippisley-Cox, J., Vinogradova, Y. and Coupland, C. (2007). Final report for the Information Centre for Health and Social Care: Time series analysis for 2001–2006 for selected clinical indicators from the Quality and Outcomes Framework. Leeds: The Information Centre.

Kontopantelis, E., Doran, T., Gravelle, H., Goudie, R., Siciliani, L. and Sutton, M. (2011). Family doctor responses to changes in incentives for influenza immunization under the Quality and Outcomes Framework pay-for-performance scheme. *Health Services Research.* In press.

ISD Scotland. (2011a) Summary of available QOF points and pounds, by year, quality and outcomes framework. Available at http://www.isdscotland.org/Health-Topics/General-Practice/Quality-And-Outcomes-Framework/QOF-Points-Pounds-Available-Summary.asp [Accessed 8 November 2011].

ISD Scotland. (2011b) List of individual QOF indicator descriptions and points values, Quality & Outcomes Framework (QOF) for April 2004 - March 2011. Available at http://www.isdscotland.org/Health-Topics/General-Practice/Quality-And-Outcomes-Framework/QOF_Scot_200405_to_200910_indicators_lookup.xls [Accessed 8 November 2011].

ISD Scotland. (2010) Primary Medical Services, the new GMS contract, and QOF. Available at http://isd.scot.nhs.uk/isd/3364.html [Accessed 12 March 2012].

Millett, C., Bottle, A., Ng, A., Curcin, V., Molokhia, M., Saena, S. and Majeed, A. (2009). Pay for performance and the quality of diabetes management in individuals with and without co-morbid medical conditions. *Journal of the Royal Society of Medicine.* 102(9), 369-377.

Morris, S., Goudie, R., Sutton, M., Gravelle, H., Elliott, R., Hole, A., Ma, A., Sibbald, B. and Skatun, D. (2011) Determinants of general practitioners' wages in England. *Health Economics*. 20(2), 147-160.

Scottish Government. (2008) General Medical Services statement of financial entitlements for 2007. Available at http://www.sehd.scot.nhs.uk/pca/PCA2008%28M%2901.pdf [Accessed 8 November 2011].

Williams, P.H. and de Lusignan, S. (2006) Does a higher 'quality points' score mean better care in stroke? An audit of general practice medical records. *Informatics in Primary Care.* 14(1), 29-40.

Shohet, C., Yelloly, J., Bingham, P. and Lyratzopoulos, G. (2007) The association between the quality of epilepsy management in primary care, general practice population deprivation status and epilepsy-related emergency hospitalisations. *Seizure.* 16(4), 351-355.

Steel, N. and Willems, S. (2010). Research learning from the UK Quality and Outcomes Framework: A review of existing research. *Quality in Primary Care*. 18(2), 117-125.

Strong, M., South, G. and Carlisle, R. (2009) The UK Quality and Outcomes Framework pay-for-performance scheme and spirometry: Rewarding quality or just quantity? A cross-sectional study in Rotherham, UK. *BMC Health Services Research.* 9, 108.

Vaghela, P., Ashworth, M., Schofield, P. and Gulliford, M.C. (2009) Population intermediate outcomes of diabetes under pay-for-performance incentives in England from 2004 to 2008. *Diabetes Care.* 32(3), 427-429.