THE UNIVERSITY *of York*

CHE
Centre For Health Economics

Doctor Behaviour under a Pay for Performance Contract: Further Evidence from the Quality and Outcomes Framework

CHE Research Paper 34

Doctor behaviour under a pay for performance contract: further evidence from the Quality and Outcomes Framework

Hugh Gravelle[a]
Matt Sutton[b]
Ada Ma[b]

National Primary Care Research and Development Centre, Centre for Health Economics, University of York, UK [a]
Health Economics Research Unit, University of Aberdeen, UK [b]

February 2008

Background

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The new CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

NPCRDC receives core funding from the Department of Health. HERU receives funding from the Chief Scientist Office of the Scottish Executive Health Department. The views expressed are those of the authors and not necessarily those of the funders. Part of the paper was written whilst HG and MS were visiting the Centre for Market and Public Organisation (CMPO), University of Bristol. We are grateful for comments made at presentations at CHE, CMPO, the Norwegian School of Economics and Business Administration, the Royal Economics Society Conference in Warwick, the International Health Economics Association Conference in Copenhagen, the Office of Manpower Economics, and departmental seminars in Dundee and Sheffield; and for discussions with Tim Doran, Bruce Guthrie, David Reeves and Martin Roland about the intricacies of the QOF. This paper is revised version of CHE RP 28 incorporating a larger number of indicators and explanatory variables.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/inst/che

Abstract

Background

Since 2003, 25% of UK general practitioners' income has been determined by the quality of their care. The 65 clinical quality indicators in this scheme (the Quality and Outcomes Framework) are in the form of ratios, with financial reward increasing linearly with the ratio between a lower and upper threshold. The numerator is the number of patients for whom an indicator is achieved and the denominator is the number of patients the practices declare are suitable for the indicator. The number declared suitable is the number of patients with the relevant condition less the number exception reported by the practice for a specified range of reasons. Exception reporting is designed to avoid harmful treatment resulting from the application of quality targets to patients for whom they were not intended. However, exception reporting also gives GPs the opportunity to exclude patients who should in fact be treated in order to achieve higher financial rewards. This is inappropriate use of exception reporting or 'gaming'. Practices can also increase income if they are below the upper threshold by reducing the number of patients declared with a condition (prevalence), or by increasing reported prevalence if they were above the upper threshold. This study examines the factors affecting delivered quality (the proportion of prevalent patients for indicators were achieved) and tests for gaming of exceptions and for prevalence reporting being responsive to financial incentives.

Data

We used routinely available data on the Quality and Outcomes Framework from Scottish practices (n=916) for 2004/05 and 2005/06. We also include data on practice characteristics and on socio-demographic and morbidity factors from the 2001 census and the Scottish Index of Multiple Deprivation.

Methods

Multiple regressions of delivered quality, exception reporting, and prevalence reports on practice and patient characteristics. We test for gaming of exception reporting by comparing the rates of exception reporting in 2005/6 for practices which were above the upper threshold in 2004/5 (which would have had no incentive to increase exception reporting) with practices which were below the threshold in 2004/5 (which would have had a financial incentive to increase exception reporting rates. We also compared prevalence reporting in 2005/6 for practices above the upper threshold in year one (who would gain financially by increasing prevalence) with those below the threshold in year one (who might not).

Results

90.8% of practices reported levels of achievement above the upper thresholds. They could have reduced the number of patients treated by 11.8% without reducing income, indicating a degree of altruistic behaviour. Delivered quality was lower in practices with more income deprived patients and with a higher proportion of ethnic minority patients, though the effects of these variables were quite small.

Practices which were above the upper threshold for an indicator in 2004/05 had higher prevalence in 2005/6 compared to practices that were below the threshold in 2004/05.

For exception reporting, practices which were below the upper threshold for an indicator in 2004/05 had higher exception reporting rates in 2005/6 than practices which were below the upper threshold. From the differences between the two types of practice, we estimate that, in practices which were below indicator thresholds in 2004/05, 0.87% of patients might have been inappropriately exception reported in 2005/06, or 10.9% of the overall number of patients exception reported.

Conclusion

The results suggest that general practitioners are partially altruistic in that the majority produced markedly higher quality than was required to maximise their financial rewards. Quality delivered to patients was higher in larger practices and in practices with less deprived populations and with a smaller ethnic minority population.

Exception reporting removes incentives towards inappropriate or over-treatment of patients. But the QOF provides perverse incentives for gaming of exceptions and we find evidence that practices which performed worse in 2004/5 were more likely to game exceptions in 2005/6.

The incentives in QOF also affect case finding and reporting by practices and we find evidence that practices which performed worse in 2004/5 had lower reported prevalence in 2005/6. We also find that reported prevalence rates are associated with practice characteristics, such as whether the practice was a fundholder, suggesting that the QOF prevalence reports may not be a reliable epidemiological resource.

Keywords. Quality. Incentives. Gaming. Pay for performance.

1. Introduction

The development of new methods of measuring the quality of health care has led to the recognition that there are wide variations in the quality of health care providers (Fisher, 2006; Institute of Medicine, 2007). It has also prompted an international interest in the use of pay for performance schemes to improve quality. Schemes are being developed in Australia (Scott, 2007; Medicare Australia, 2007), Canada (Pink et al, 2006), Germany (Greb et al, 2006), The Netherlands (Custers T et al, 2006) New Zealand (Perkins and Seddon, 2006), Spain (Gené-Badia et al, 2007), the US (Rosenthal, 2005, 2006; Epstein, 2006; Institute of Medicine, 2006), and the UK (Roland, 2004).

The Quality and Outcomes Framework (QOF) introduced in 2004 in the UK National Health Service (NHS) is possibly the most elaborate and expensive health care pay for performance scheme. All 10,000 general (primary care) practices in the UK were required to report their achievements on 146 quality indicators. The average practice, containing four general practitioners (GPs), stood to gain around £130,000 per year in 2005/6 if it achieved all indicators to the maximum extent. The amount paid out to practices under the QOF was around £1,000M in 2005/6.[1]

The QOF was a good deal for GPs. It increased their income by about 25% (£20,000 per year), reduced their hours of work, and increased their job satisfaction (Whalley, Gravelle and Sibbald, 2007). The effect on patients is less clear. The tentative conclusion, from studies based on subsets of the indicators (Hippisely-Cox, Vinogradova, and Coupland, 2007; Gulliford et al, 2007; Campbell et al, 2007; Steel et al, 2007; Tahrani et al, 2007), is that the quality of care had been increasing since the late 1990s, but that the QOF may have led to above trend increases in some of the activities it incentivises.

One of the difficulties in designing P4P schemes linking financial rewards to measures of performance is that the incentives may generate unanticipated and dysfunctional behaviour (Prendergarst, 1999). In this paper, we use some of the features of the QOF, in particular a sharp discontinuity in the link between reward and performance, to test for gaming. We show that a scheme based on ratio quality measures, intended to incentivise doctors to increase the numerator also contained incentives to manipulate the denominators, and that some doctors did so.[2]

Around half of potential QOF revenue is attached to indicators of clinical quality measured as the ratio of treated patients to the number of patients reported to be eligible for the indicator. For example, indicator CHD6 is the proportion of eligible patients with coronary heart disease whose blood pressure is controlled. QOF revenue increase linearly with the proportion treated, between a lower threshold (0.25) and an upper threshold (0.70) (see Figure 1 and Table 1).

Table 1. Example of indicators from Coronary Heart Disease domain

Indicator	Max points	Upper threshold
CHD1: Practice has register of patients with CHD (yes/no)	6	
CHD5: Percentage of CHD patients whose notes record BP in previous 15 months	7	90%
CHD6: Percentage of CHD patients whose BP in previous 15 months is 150/90 or less	19	70%
CHD11: Percentage of patients with a history of myocardial infarction (diagnosed after 1 April 2003) who are currently treated with an ACE inhibitor	7	70%

All lower thresholds in all disease domains are 25%

[1] 2005/6 is the financial year 1 April 2005 to 31 March 2006. Similarly for the financial year 2004/5.
[2] We briefly report on the factors affecting practice points achievement and quality in Appendix D.

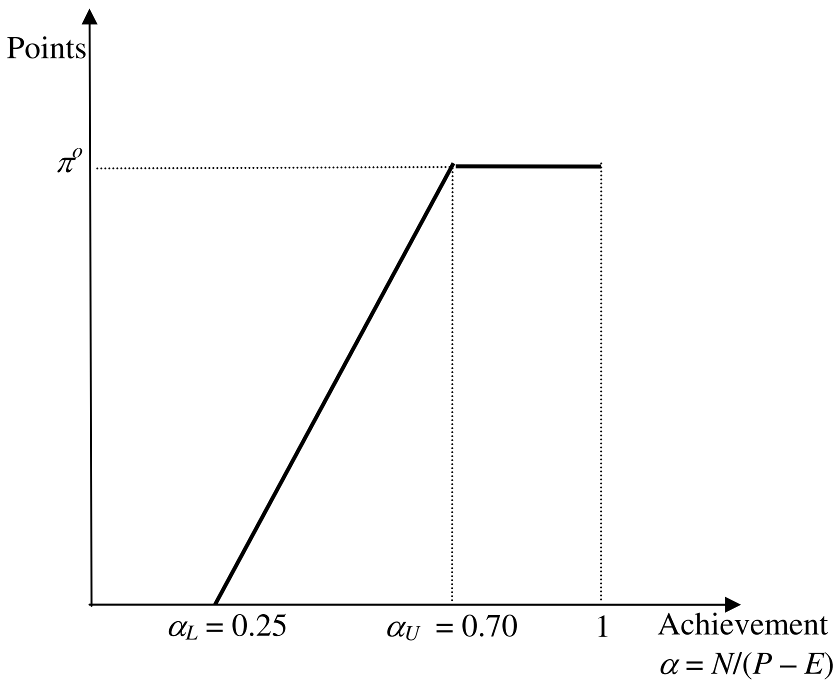

Figure 1. Achievement and points for a continuous quality indicator (CHD6)

Expressing a clinical indicator as a ratio is intended to provide an incentive to practices to increase the numerator, i.e. to increase the number of treated patients. But practices can also reduce the denominator, i.e. the number they declare as eligible for treatment. The number declared eligible is the number of patients the practice reports as having a condition (prevalence) minus the number it exception reports for the indicator. Patients can be reported as unsuitable for an indicator on a variety of grounds (Appendix A), for example because they are terminally ill or have failed to attend for treatment. Thus practices can increase QOF revenue by gaming their reports of both prevalence and exceptions.

We test for gaming of prevalence and exception reporting in two ways. First, true disease prevalence and exception reporting rates are determined by the characteristics of the population served by the practice. They should not vary with characteristics of the practice, such as the type of contract it holds or the number of neighbouring practices. We find that prevalence and exception reporting rates vary with population characteristics, such as ethnicity and deprivation, in expected ways. But reporting was also significantly affected by characteristics of the practice. For example, exception reporting rates were higher in practices where patients had fewer practices to choose from. This test for gaming is vulnerable to the omission of population characteristics that affect true prevalence or exceptions and are correlated with the included practice characteristics. We attempt to reduce such omitted variable bias by including measures of practice population morbidity, deprivation, rurality, and ethnicity.

Our second test uses the structure of the incentive scheme. The marginal reward for overstating exceptions is positive when the quality ratio is below the upper threshold and zero when it is above the upper threshold. Similarly, practice revenue is increased by understating prevalence when the quality ratio is below the upper threshold. Moreover, because of another feature of the scheme explained below, practice revenue is increased by overstating prevalence when the quality ratio is above the upper threshold.

The price per point increased by 75% between the first and second years of the QOF. Thus, for practices below the upper threshold for an indicator there was a considerable increase in the marginal financial reward for increasing the proportion of eligible patients treated. For practices above the threshold the marginal financial reward from increasing the quality ratio treated was zero in both years. We argue that practices which were below the upper threshold in the first year of the scheme would have higher reported exception rates and lower reported prevalence in the second year than practices above the upper threshold. Practices above the upper threshold in 2004/5 would only have to maintain their behaviour to maximise their 2005/6 revenue from the QOF. We find that practices that were below the upper threshold in the first year had higher exception rates and lower prevalence rates in the second year than practices that were above the threshold in the first year. Practices below the upper threshold in 2004/5 had an average exception rate of 8.55% in 2005/6 and we estimate that without the incentive to increase their exceptions they would have had an exception rate of 7.25 %. Practices which were above the upper threshold in 2004/5 had reported standardised prevalence in 2005/6 which was 3.5% greater, against a mean of 100, than if they had been below the upper threshold in 2004/5.

Section 2 describes the incentives for family doctors in the QOF. Section 3 specifies a simple model of the behaviour of semi-altruistic family doctors under the QOF. It shows how the discontinuity in the link between payment and quality provides a means of testing for gaming. The data are described in Section 4 and results are presented in Section 5. Section 6 discusses the implications of the results for the design of pay for performance schemes.

2. Incentives in the QOF

2.1 Family doctor contracts

The NHS is financed almost entirely from general taxation. Patients register with general practices, which act as gatekeepers for elective hospital care. Patients face almost no charges for NHS health care.[3]

GPs are independent contractors, apart from a small minority (0.8% in Scotland) who are directly employed by their local primary care organisation, They are organised in partnerships with a mean size of 4.4 GPs and around 5,200 patients.

GPs are paid under one of two contracts. Most GPs (90% in Scotland) are in practices with a nationally negotiated General Medical Services (GMS) contract under which they are paid by a mixture of capitation, lump sum allowances, items of service, and target incentives including the QOF. The capitation payments vary with the age of patients and with the deprivation level of the area in which they lived. GPs have to meet all their practice expenses from their gross income, except for some specific reimbursements for the costs of practice nurses and computing systems. Additionally, where there is no local pharmacy, GPs are permitted to dispense the medicines they prescribe. Dispensing practices can make a profit from dispensing since they receive a dispensing fee per item and are reimbursed for the drugs they buy at a rate that often exceeds the price they paid.

Around a tenth of Scottish practices have opted to be paid under a Primary Medical Services (PMS) contract. These contracts are negotiated between the practice and their local primary care organisation (Health Board in Scotland). Under the PMS contract, the practice receives a lump sum in exchange for agreeing to provide the services they would have provided under the GMS contract, plus additional services for particular patient groups. The amount received is typically the amount the practice would have received under GMS, plus an addition intended to cover the cost of the extra services. As under GMS, the practice has to meet its expenses from its gross income. PMS practices were required to take part in the QOF, but, because it was thought that they would already be being paid for some of the services counting towards the QOF, they had points deducted from their QOF score.[4]

2.2 Quality indicators and practice revenue

The QOF rewards practices according to quality indicators for four areas: clinical (covering eleven chronic diseases), organisation (for example whether the practice has an annual review of patient complaints and suggestions), patient experience (for example whether the practice has undertaken an approved patient survey in the past year), and additional services (the practice offers ante natal care and screening). (See Tables 2,3) There are additional points for holistic care (determined by the

[3] The only charge is for drugs prescribed in primary care. Because of the wide range of exemptions on grounds of age, income, and health, 92% of drug prescriptions carried no charge in 2005/6.
http://www.isdscotland.org/isd/info3.jsp;jsessionid=A436EB8C9E9ED916E8D06A4DDB99020C?pContentID=2237&p_applic=CCC&p_service=Content.show&

[4] The deduction was168 points in 2004/5 and 109 points in 2005/6, from a maximum possible score of 1050 points (www.bma.org.uk/ap.nsf/Content/pmsagreements0904~pmsnewgms?OpenDocuments&Highlight=2,PMS,QOF)

percentage of the available points scored on the third worst disease domain in the clinical quality area), for quality practice (determined by the third worst performance in the domains of the three non-clinical areas), and a bonus based on an access survey.[5]

Table 2. QOF areas: indicators and points

	Indicators	Maximum points
Clinical quality	76	550
Organisation	56	184
Patient experience	4	100
Additional services	10	36
Holistic care		100
Quality practice		30
Access bonus		50
Total	146	1050

We concentrate on the 65 ratio clinical indicators that are measured as the proportion of relevant patients for whom an indicator has been achieved. For example, indicator CHD5 is the proportion of patients with Coronary Heart Disease whose notes have a record of blood pressure in the last 15 months. The indicators are listed in Table 3. The indicators carry points that have a monetary value which varies with the size of the practice and with the proportion of the practice population in each disease domain.

Achievement by a practice on a ratio clinical quality indicator is $\alpha = N/D = N/(P-E)$. N is the number of patients for whom the indicator is achieved eg the number with CHD whose blood pressure has been measured in the last 15 months. $D = P - E$ is the number of patients declared as suitable for the indicator. Prevalence P is the number of patients in the disease domain reported by the practice. E is the number of patients who are exception reported by the practice because the indicator is not appropriate for them. Patients may be exception reported on nine grounds (Appendix A). For example, terminal cancer patients with hypertension would be excluded from the hypertension indicators. Patients who have been invited to attend for treatment on three occasions in the preceding twelve months and have failed to attend can also be excluded.

The number of points (and hence revenue) earned on an indicator varies linearly with achievement for $\alpha \in (\alpha_L, \alpha_U)$. See Figure 1. No points are earned if α is less than or equal to the lower threshold α, which is 25% for all indicators. At the upper threshold α the practice receives the maximum number of points π^o available for the indicator and further increases in achievement have no effect on revenue. The upper thresholds vary across indicators, from 50% to 90%, and the maximum points vary from 2 to 56 (Table 3).[6]

Formally, the number of points earned for an indicator is

$$\pi = \pi(\alpha) = \pi^o \min\{1, \max\{(\alpha - \alpha_L)/(\alpha_U - \alpha_L), 0\}\} \qquad (1)$$

The revenue from the indicator is the number of points times the price per point. The price per point is the product of the national average price per point v, the list size of the practice relative to the national average list size $M/\bar{M}$, and the adjusted practice disease prevalence factor F for the disease domain containing the indicator:

$$R = \pi(\alpha) v \frac{M}{\bar{M}} F(P) \qquad (2)$$

The national average price per point v was £70 in 2004/5 and £124.60 in 2005/6.

[5] The large number of indicators and the holistic payment are an attempt to reduce effort diversion problems (Holmstrom and Milgrom, 1991) arising when agents perform several socially valuable tasks.

[6] Setting upper thresholds below 100% was intended to reduce the risk that GPs would be inappropriately treat some patients, for example those with co-morbidities. Exception reporting was justified on the same grounds. The rationale is peculiar for those indicators where there is no risk to patients, for example the recording of blood pressure, as opposed to providing treatment for hypertension. Upper thresholds below 100% make it easier to earn points and increase the marginal reward for treating a patient.

Table 3. Disease domains: ratio clinical quality indicators, achievement, and reported prevalence. Scottish practices 2004/5, 2005/6.

Disease domain	Number of indicators	Points available	Upper threshold achievement 2004/05		Upper threshold achievement 2005/06		Reported raw prevalence rate (%) 2005/06	
			Mean	SD	Mean	SD	Mean	SD
Asthma	6	65	0.75	0.26	0.94	0.15	5.44	1.09
Cancer	1	6	0.76	0.39	0.87	0.31	0.72	0.23
CHD	11	95	0.81	0.22	0.95	0.10	4.58	1.05
COPD	7	40	0.70	0.30	0.89	0.19	1.86	0.92
Diabetes	17	93	0.81	0.19	0.91	0.12	3.40	0.72
Epilepsy	3	14	0.65	0.30	0.84	0.24	0.71	0.21
Hypertension	4	96	0.72	0.35	0.92	0.20	12.07	2.78
Hypothyroidism	1	6	0.93	0.21	0.98	0.13	2.96	0.92
LVD	2	16	0.76	0.38	0.88	0.31	0.65	0.26
Mental health	4	34	0.76	0.38	0.88	0.31	0.60	0.33
Stroke and TIA	9	27	0.78	0.24	0.95	0.12	1.92	0.54
Total	65	492						

Prevalence rate: List weighted mean of raw prevalence rates.
Upper threshold achievement is the unweighted average across practices of *Domain upper*. *Domain upper* is the average proportion of indicators in a domain for which the practice achieved the upper threshold, weighted by maximum possible points for the indicator.
CHD: coronary heart disease. COPD: chronic obstructive pulmonary disease. LVD: left ventricular dysfunction. TIA: transient ischaemic attacks.

The adjusted disease prevalence factor F is the square root of the practice disease prevalence rate divided by the unweighted average of the square roots of the practice prevalence rates in all practices:[7]

$$F = F(P) = \left(P/M\right)^{\frac{1}{2}}\left[G^{-1}\sum_h^G\left(P_h/M_h\right)^{\frac{1}{2}}\right]^{-1}, \qquad F'(P) > 0 \tag{3}$$

Where G is the total number of practices. Thus practice revenue from an indicator is[8]

$$R = v\left[M/\bar{M}\right]\pi(\alpha)\left[P/M\right]^{\frac{1}{2}}\left[G^{-1}\sum_h(P_h/M_h)^{\frac{1}{2}}\right]^{-1} \tag{4}$$

A practice of given list size can alter its QOF revenue by changing

- the number of patients for whom the indicator is met (N)
- the number of patients who are exception reported for that indicator (E)
- the reported number of patients with the relevant condition (P)

When $\alpha \in (\alpha_L, \alpha_U)$ increasing the number of patients for whom the indicator is met (N) or increasing exception reporting (E) will increase revenue.

The effect of reporting a larger number of prevalent patients (P) is more complicated. An increase in P reduces achievement $\alpha = N/(P - E)$ and increases the adjusted disease prevalence factor $F(P)$. If $\alpha < \alpha_L$ then $\pi'(\alpha) = 0$ and $\pi(\alpha) = 0$, so that $\partial R/\partial P = 0$. When $\alpha > \alpha_U$ we have $\pi'(\alpha) = 0$ and $\pi(\alpha) > 0$, so that $\partial R/\partial P > 0$ because of the increase in the adjusted disease prevalence factor. In the intermediate range $\alpha \in (\alpha_L, \alpha_U)$ increases in P reduce achievement α but increase the adjusted disease prevalence factor. We can show (Appendix B) that the overall effect is to reduce revenue from the indicator. Thus

$$\begin{aligned}\frac{\partial R}{\partial P} &= \bar{v}\pi'(\alpha)\frac{\partial\alpha}{\partial P}F(P) + \bar{v}\pi(\alpha)F'(P) \\ &= 0, \qquad \alpha < \alpha_L \\ &< 0, \qquad \alpha \in (\alpha_L, \alpha_U) \\ &> 0, \qquad \alpha > \alpha_U\end{aligned} \tag{5}$$

with marginal revenue jumping down from zero at α_L and jumping from negative to positive at α_U.

---

7 We ignore an additional complication that the practice prevalence rate should have been truncated at the 5th centile both in calculating the practice rate and the national average rate. It appears that in Scotland the truncation was applied only to the very few practices with prevalence less than 5% of the national mean. The effect of the square root transformation is to reduce the difference in rewards arising from differences in prevalence rates. Guthrie, McLean and Sutton (2006) investigate its effect on practice revenue.

8 Because prevalence rates vary across disease domains and practices and the upper threshold varies across indicators the full formula for revenue for indicator i in disease domain k for practice g is

$$R_{gki} = v\left[M_g/\bar{M}\right]\pi_{ki}(\alpha_{gki})\left[P_{gk}/M_g\right]^{\frac{1}{2}}\left[G^{-1}\sum_h(P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1}$$

where $\pi_{gki} = \pi_{ki}(\alpha_{gki}) = \pi_{ki}^o \min\{1, \max\{(\alpha_{gki} - \alpha_{kiL})/(\alpha_{kiU} - \alpha_{kiL}), 0\}\}$ and $\alpha_{gki} = N_{gki}/(P_{gk} - E_{gki})$. A change in reported prevalence P_{gk} affects all the indicators in disease domain k, so the effect on practice revenue is $\sum_i \partial R_{gki}/\partial P_{gk}$ which depends on the mix of indicators in the domain for which the proportion of patients for whom the indicator is achieved is above or below the upper threshold.

3 Practice behaviour under the QOF

3.1 Optimal treating and cheating

We now specify a model of practice behaviour under the QOF. We assume that GPs are quasi-altruistic: they care about their income and about patient welfare (McGuire, 2000). Practice utility is linear in patient health and practice income (QOF revenue minus costs) and there is only one quality indicator:

$$\begin{aligned} s(N,E,P) &= bN + R(N,E,P) - C(N,E,P) \\ &= bN + \bar{v}\pi(\alpha)F(P) - c_{N1}N - c_{E1}\left(\left|E-E^o\right|/P^o\right) - c_{P1}\left(\left|P-P^o\right|/M\right) \\ &\quad -\tfrac{1}{2}\left[c_{N2}N^2 + c_{E2}\left((E-E^o)/P^o\right)^2 + c_{P2}\left((P-P^o)/M\right)^2\right] \end{aligned} \tag{6}$$

where $\alpha = N/(P-E)$ is practice achievement and $\bar{v} = vM/\bar{M}$. Patient health is proportional to the number treated and the parameter $b \geq 0$ reflects both technology (the effect of treatment on patient health) and the strength of GP altruism.

E^o/P^o is the exception rate if the practice used the exception criteria properly and did not game. $E/P - E^o/P^o$ is the rate of exception gaming. Similarly P^o/M is the true prevalence rate and $P/M - P^o/M$ is the rate of gaming of prevalence reporting. If there was no financial or psychic cost to gaming then all practices would achieve at least the upper thresholds and score maximum points. The inclusion of $E/P - E^o/P^o$ and $P/M - P^o/M$ in the cost function is a means of capturing the idea that gaming has a cost which may be a psychic cost of offending against professional ethics or the certainty equivalent of a financial penalty if caught cheating. The cost parameters (c_N, c_{E2}, c_{P2} are all positive and c_{E1}, c_{P1} are non-negative. The cost function derivatives with respect to E and P jump from negative to positive at E^o and P^o.

The derivatives of s with respect to numbers treated, exceptions, and prevalence are[9]

$$\frac{\partial s}{\partial N} = b + \bar{v}\pi'(\alpha)\frac{1}{P-E}F(P) - c_{N1} - c_{N2}N \tag{7}$$

$$\frac{\partial s}{\partial E} = \bar{v}\pi'(\alpha)\frac{N}{(P-E)^2}F(P) - c_{E1}\frac{d\left|E-E^o\right|}{dE}\frac{1}{P^o} - c_{E2}\frac{(E-E^o)}{P^{o2}} \tag{8}$$

$$\begin{aligned} \frac{\partial s}{\partial P} &= -\bar{v}\pi'(\alpha)\frac{N}{(P-E)^2}F(P) + \bar{v}\pi(\alpha)F'(P) \\ &\quad - c_{P1}\frac{d\left|P-P^o\right|}{dP}\frac{1}{M} - c_{P2}\frac{(P-P^o)}{M^2} \end{aligned} \tag{9}$$

The seven possible solutions to the problem of maximising **Error! Reference source not found.** subject to non-negativity constraints on N, E, and P and to $\alpha = N/(P-E) \in [0,1]$ are illustrated in Figure 2.[10]

(i) $\alpha = 0$. With sufficiently low altruism and high marginal costs of treatment and gaming the practice provides no treatment and does not game: $N = 0$, $E = E^o$, $P = P^o$.

(ii) $\alpha \in (0, \alpha_L)$. The practice's altruism is sufficiently great and marginal cost of treatment sufficiently small to lead it to treat some patients but not enough to achieve the lower threshold.

[9] Except at the lower threshold where $\pi'(\alpha)$ jumps from zero to become positive and at the upper threshold were it jumps down to zero and at $E = E^o$ (and $P = P^o$) where there are upward jumps in the cost function with respect to E (and P).

[10] Preferences in R, N space are given by $S(R,N;E,P) = R + bN - C(N,E,P)$ with the revenue function $R = R(N;E,P)$ given by **Error! Reference source not found.**. The marginal utility of treatment is $S_N = b - c_N N$.

Because it has not achieved the lower threshold there is no income gain from gaming exceptions or prevalence. Hence $P = P^o$, $E = E^o$ and $\alpha_L(P^o - E^o) > N > 0$.

(iii) $\alpha = \alpha_L$. This cannot be a solution. If $\alpha = \alpha_L$ is the solution it must be true that $\partial s / \partial N$ is non-negative for $\alpha < \alpha_L$ and non-positive for $\alpha > \alpha_L$. But the derivative of *u* with respect to *N* jumps upward at $\alpha = \alpha_L$.

(iv) $\alpha \in (\alpha_L, \alpha_U)$. The practice fails to achieve the upper threshold but it may game by overstating exceptions $E > E^o$ and understating prevalence $P < P^o$ if c_{E1} and c_{P1} are small enough.

(v) $\alpha = \alpha_U$. The practice just achieves the upper threshold. It may do so by overstating exceptions and understating prevalence if c_{E1} and c_{P1} are small enough.

(vi) $\alpha \in (\alpha_U, 1)$. The practice has high enough altruism and low enough marginal cost of treatment that it is willing to treat patients even though the marginal treated patient reduces practice income. Since the marginal revenue from increased achievement is zero it does not game exceptions ($E = E^o$) even if $c_{E1} = 0$ since this will merely increase its costs. Revenue is increased by higher reported prevalence and the practice will overstate prevalence $P > P^o$ if c_{P1} is sufficiently small.

(vii) $\alpha = 1$. If $N/(P^o - E^o) = 1$ marginal revenue from reported prevalence is positive. But if c_{P1} is sufficiently large there is no incentive to overstate prevalence even at $P = P^o$. Thus the solution has $E = E^o$, $P \geq P^o$ and $N = P - E^o$.

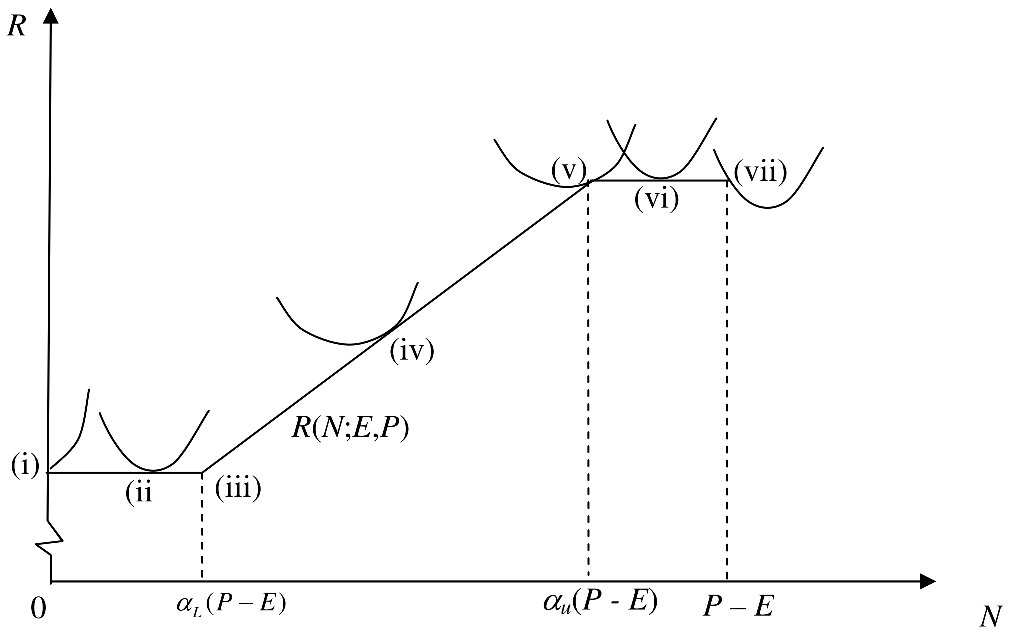

Figure 2. Solution types

The solution types are summarised in Figure 3. The optimal values of *N*, *E*, and *P* depend on the practice's altruism parameter *b*, cost parameters, true prevalence P^o, true exceptions E^o, and list size *M*. The utility function *s*(*N*,*E*,*P*) is not concave and has discontinuous derivatives at the upper and lower thresholds. Nevertheless, it is possible to derive some comparative static predictions (see Appendix C for details). We can establish

Proposition 1. For any given vector of cost parameters, true prevalence, true exceptions and list size, there exist b_1, b_2, b_3, b_4 (with $b_1 < b_2 < b_3 < b_4$) such that for all $b < b_1$ the practice optimum has $\alpha < \alpha_L$, for $b \in (b_2, b_3)$ the optimum has $\alpha \in (\alpha_L, \alpha_U)$ and for $b > b_4$ the optimum has $\alpha > \alpha_U$.

It can also be shown that, holding all other parameters constant, changes in the cost function parameters which reduce the marginal cost of treatment will also shift the optimum in the same way as the increase as the altruism parameter. We use the proposition to derive tests for gaming of prevalence and exceptions.

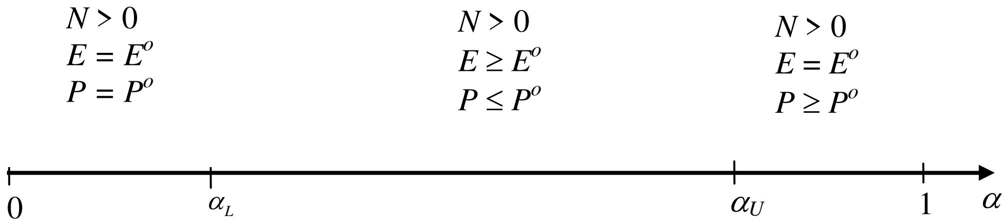

Figure 3. Solution types and achievement level

3.2 Testing for gaming

Prevalence reporting below and above the upper threshold

Very few practices had $\alpha < \alpha_L$ for any indicator. We therefore test for gaming by comparing practices with $\alpha \in (\alpha_L, \alpha_U)$ and $\alpha > \alpha_U$. The marginal revenue from an increase in declared prevalence (P) is negative for practices with $\alpha \in (\alpha_L, \alpha_U)$ and positive for practices with $\alpha > \alpha_U$. Holding all other characteristics constant, practices with $b > b_4$ (or low enough marginal cost of treatment) will have $\alpha > \alpha_U$ and hence will have declared prevalence rates P/M which are greater than those with $b \in (b_2, b_3)$ (or high enough marginal treatment cost). This suggests that a comparison of the reported prevalence rates of practices above and below the upper threshold, conditioning on observable practice and patient characteristics affecting cost parameters, true prevalence, and true exceptions, provides a test for gaming of prevalence reporting.

But a test for gaming of prevalence by comparison of practices above and below the upper threshold faces the obvious difficulty that unobservable factors that directly increase reported prevalence (for example sicker patients) will make it less likely that a practice with given observable characteristics will be above the threshold. Thus the estimated effect of being above the upper threshold would be biased downwards and could be negative even if practices were indeed gaming.

Since the new 2004/5 contract embodying the QOF was a radical departure from the previous contract it is plausible that practices used their achieved performance in 2004/5 to inform decisions affecting 2005/6 performance. Those below the upper threshold would realise that they needed to increase treatment (N) or reduce reported prevalence (P) or increase exceptions (E) to generate additional QOF revenue in 2005/6. Those above the upper threshold would know that they could only increase QOF revenue by increasing prevalence. Moreover, the average price per point increased by 75% between 2004/5 and 2005/6. This increased the marginal rewards for reducing prevalence reporting for practices expecting to be under the threshold in 2005/6 and increased the marginal rewards for increasing reported prevalence for those expecting to be above the threshold in 2005/6.

We therefore test for gaming of prevalence reporting by examining whether practices which were above the upper threshold in 2004/5 had a higher reported prevalence in 2005/6 than those which were above the upper threshold in 2004/5. We regress reported prevalence in 2005/6 on a set of practice and patient characteristics to capture differences in costs, true prevalence and true exceptions, and a dummy $U_{04/05}$ indicating whether the practice was above the upper threshold in 2004/5.

A positive coefficient on $U_{04/05}$ is evidence for gaming provided that the unobservable factors affecting reported prevalence in 2005/5 are not positively correlated with $U_{04/05}$, conditional on the patient and practice covariates. This is certainly the case if the unobservables affecting the 2005/6 reported prevalence rate, and the 2004/5 prevalence rate, numbers treated, and numbers excepted are uncorrelated.

If the unobservables affecting 2004/5 and 2005/6 prevalence rates are positively correlated then, in the absence of gaming, the coefficient on $U_{04/05}$ will be negative. Hence a positive coefficient on $U_{04/05}$

would be even stronger evidence of gaming. It is highly implausible that the unobservables affecting 2004/5 and 2005/6 prevalence rates are negatively correlated.

Even if there are persistent unobservable factors (for example some aspect of patient deprivation not captured in our deprivation measures) which increase prevalence and numbers treated, this does necessarily induce a positive correlation of $U_{04/05}$ and the 2005/6 reported prevalence rate. An increase in N and P will increase achievement ($N/(P-E)$) only if the ratio of the increases in N and P exceeds $N/(P-E)$. Similarly persistent factors increasing E and P only induce a positive correlation of $U_{04/05}$ and the 2005/6 reported prevalence rate if they have a larger effect on E than P. For example, even though a practice with more unobserved deprivation might have a higher prevalence rate and be able to legitimately exception report more patients, this would induce a positive coefficient on $U_{04/05}$ only if unobserved deprivation had a bigger effect on the probability that a prevalent patient is legitimately exception reported than on the prevalence rate.

Since we also include a rich set of patient and practice characteristics (see section 4) in the prevalence regression model, and also run further models with practice fixed effects, we argue that positive coefficients on $U_{04/05}$ from such regressions are evidence for gaming of prevalence.

Exception reporting above and below the upper threshold

We use a similar method of testing for gaming of exception reporting. Since marginal revenue from increased exception reporting is positive when $\alpha \in (\alpha_L, \alpha_U)$ and zero when $\alpha > \alpha_U$, practices with low enough costs of gaming exceptions will have more exceptions when their altruism parameter is low enough (or marginal cost of treatment is high enough) to ensure that they choose $\alpha \in (\alpha_L, \alpha_U)$ than when it is high enough that they choose $\alpha > \alpha_U$. The argument also implies that exceptions as a proportion of true prevalence P^o will be higher in practices below the threshold. Since such practices will also have a lower declared prevalence, exceptions as a proportion of declared prevalence will also be higher for practices below the upper threshold.

As with prevalence reporting, we cannot test for gaming by a regression of exception reporting rates in 2005/6 on $U_{05/06}$ since unobserved factors increasing exceptions in 2005/6 will increase the likelihood that a practice is above the upper threshold in 2005/6. Hence the test for gaming is biased towards rejection of gaming. We therefore test for exception report gaming by regressing the 2005/6 exception rate on practice and patient characteristics and the upper threshold indicator for 2004/5. Gaming would lead to higher exception reporting in 2005/6 for practices below the threshold in 2004/5 compared to those above the upper threshold: the coefficient on $U_{04/05}$ will be negative.

As with the test of prevalence we argue that whilst it is not impossible that persistent unobservable factors increasing 2005/6 exceptions and reducing 2004/5 achievement could also explain a negative coefficient on $U_{04/05}$, the use of a rich set of practice and patient covariates makes it unlikely. We also use practice effects to remove such persistent factors.

Effects of patient and practice characteristics on reported prevalence and exceptions

P^o and E^o (true prevalence and exceptions) depend on the characteristics of the practice population, not on the characteristics of the practice such as the age of GPs, their country of qualification or the type of practice contract. Gamed prevalence and exceptions will vary with practice characteristics that reflect GP cost or preference parameters and with patient characteristics. Thus if practice characteristics are associated with reported prevalence and exception reporting, this suggests that practices are gaming exceptions, provided there are no omitted practice patient characteristics which affect true prevalence and exceptions and are correlated with included practice characteristics. We include measures of practice population demographics, population morbidity, deprivation, rurality, and ethnicity to reduce this omitted variable problem.

4. Data and methods

4.1 Data: covariates

Practice characteristics

Practice characteristics were taken from the GP Contractor Database held at the Information Services Division (ISD) of the Scottish Executive and are for 1 April 2005. They include practice list size, and the number, gender, ages and country of qualification of practice GPs (see Table 5).

Before the introduction of the QOF, practices with a PMS contract (section 2.1) had a higher income than those with the standard General Medical Services contract. It is possible that they would therefore put less effort into QOF activities because of income effects. We therefore include an indicator *PMS contract* in the regression models.

Most patients are required to take prescriptions to a community pharmacy to have the prescribed items dispensed. Where there is no community pharmacy available locally, practices are allowed to both prescribe and dispense items to patients. Dispensing practices are paid dispensing fees and are reimbursed for drugs dispensed at a rate which can exceed their actual cost to the practice. Dispensing practices have higher workload and income, and hence may respond less to the incentive effects of the QOF. We classified practices allowed to dispense directly to at least one of their patients as dispensing practices and include this as a covariate (*Dispensing*). Dispensing status is more common in more remote areas, but we also include a measure of rurality based on the size of local towns (*Rurality*) and the inverse of the population density (*Sparsity*) amongst the population covariates.

Qualified GPs can apply to be trainers of newly qualified doctors intending to become GPs. We include GP trainers may provide better quality care This may indicate higher quality practice. But, trainee doctors may provide lower quality care. We therefore include a dummy variable *Training* indicating whether a practice has training status.

Between 1991/2 and 1998/9 practices could opt to hold a budget to cover elective inpatient costs and drug costs for their patients (Dusheiko et al, 2006) and could retain any budget surplus to spend for the benefit of their patients. Since fundholding scheme was voluntary, we use the practice's participation in it as an indication that it was more alert to the financial implications of the QOF. To allow for the creation of new practices since the end of fundholding we classify practices as new practices (*New practice*) if they were not in existence in 1999, and those that were in existence in 1999 either as ex fundholders (*Ex FH*) or non-fundholders. Non-fundholders are the baseline category in the regressions.

We constructed measure of the extent to which each practice faced potential competition from other practices. Scotland contains 6505 small areas (datazones) containing between 500 and 1000 residents. We first calculated the Herfindahl index for each datazone based on the squared shares of the datazone population on the lists of different practices and then took a weighted average of the datazone Herfindahls where the weights were the proportion of the practice population drawn from each datazone. Since practices are paid a capitation payment for each patient registered with them, we expect practices to compete on the basis of quality and therefore that the *Herfindahl* index will be negatively correlated with quality.

Around one third of practices in Scotland are part of the Scottish Programme for Improving Clinical Effectiveness in Primary Care (SPICE-PC) network. The network provides them every six with feedback on their relative performance on a wide range of indicators, some of which are included in the QOF. The *SPICE* dummy equals 1 if the practice participated in the Autumn 2004 round of SPICE. Since the data for SPICE are extracted from computerised patient records, we expect SPICE participants to have had more comprehensive records prior to the introduction of the QOF and therefore higher prevalence.

Practices have a choice of patient record computer systems. One of systems (GPASS) is subsidised by central government. The functionality of GPASS has been criticised and alternatives are perceived to better enable practices to manage their QOF activities. We include information on the five types of computer system in use by practices as at March 2005.

Population characteristics

We use three measures of population health in the local area that we expect to be predictive of true prevalence. The Standardised Illness Ratio (*SIR*) is an indirectly age standardised measure of the proportion of people reporting a limiting long-term illness at the 2001 Census. Census data are available for each of 42,604 Census Output Areas containing on average 117 individuals in private households. The Census also has information on self assessed health in three categories ('good', 'fairly good' and 'not good'. We also calculated the Standardised Not Good Health Ratio for each Output Area as the indirectly standardised proportion of people reporting that their general health was 'not good'. It was highly collinear with SIR and we do not report results from the regressions containing it.

Our third measure of population health is the Standardised Mortality Ratio (*SMR*), an indirectly standardised measure calculated by ISD for each data zone using General Register Office death records for 1999 to 2003 and the 2001 Census population counts.

The Census provided the proportion of residents in black and ethnic minority groups in each Output Area. Measures of the socio-economic status of practice populations are derived from the Scottish Index of Multiple Deprivation 2004 (www.scotland.gov.uk/stats/simd2004/). We use the scores for six deprivation domains: income; education; employment; housing; skills and training; and access to services. The scores are available at data zone level. The ethnicity and deprivation variables are expected to affect prevalence and may influence some of the reasons for exceptions.

The practice values of the measures of health, deprivation, and ethnicity variables were calculated as the average of the variable for each geographical area (output area or data zone) from which the practice draws its population weighted by the share of practice population in each geographical area.

4.2 Measurement of QOF variables

Data on practice achievement (*N*), numbers declared suitable for an indicator (*D*) and declared prevalence (*P*) for Scottish practices for 2004/5 and 2005/6 are from ISD (www.isdscotland.org/QOF). Only practice-level counts are available.

We estimate models to test for gaming of exceptions and prevalence reporting. For 2005/6 there are data on the numbers achieved for each indicator *i* in each disease domain *k* in each practice (N_{gki}), the numbers declared eligible for the indicator (D_{gki}) and the number exception reported (E_{gki}) as at 31 March 2006. But, to speed up the payment of practices, reported prevalence for a disease domain P_{gk} is recorded on 14th February, six weeks before the end of the financial year.

Exception and prevalence measures
The relationship between exceptions and numbers declared for indicator *i* in domain *k* by practice *g* is

$$D_{gki} + E_{gki} = P_{gk} + A_{gk} = P_{gk}^{T} \tag{1}$$

where A_{gk} is the number of additional patients found by the practice in disease domain *k* between 15th February and 31st March. At 31st March the total number of patients the practice could potentially declare for the denominator of the indicator is the number of patients in the disease domain declared as prevalent on 14th February plus new patients recorded between 15th February and 31st March as having the disease: $P_{gk}^{T} = P_{gk} + A_{gk}$. Although A_{gk} is not recorded we do know D_{gki}, E_{gki}, and P_{gk} for 2005/6 and hence can calculate $P_{gk}^{T} = D_{gki} + E_{gki}$. We measure the exception reporting rate for indicator *i* in disease domain *k* as $E_{gki}/ (D_{gki} + E_{gki})$.

We expect that age and sex mix of the practice population will affect reported prevalence P_{gk} since demography is the main determinant of true prevalence. We could allow for demographics by including the age and gender proportions of the practice population as explanatory variables. This would have a cost in terms of degrees of freedom and would assume no interaction between the effects of demography and other variables. We have instead used the indirectly standardised reported prevalence ratio as the dependent variable in the regressions investigating reported prevalence.

Expected prevalence figures for each practice were obtained by applying age and sex specific prevalence rates, obtained directly from practice records in a reference sample of 44 practices, to all practice populations.[11] The indirectly standardised reported prevalence ratio was obtained by dividing the number of prevalent patients reported on 14 February 2005 (P_{gk}) by the expected number $\hat{P}_{gk}$ and multiplying by 100.[12] We report prevalence regressions for 2005/6 in eight of the eleven disease domains because of the lack of information on age and sex specific prevalence rates in cancer, mental health, and LVD. The correlation between raw reported prevalence rates and indirectly standardised prevalence ratios for 2005/6 across all practices and the eight disease domains was 0.83 (p<0.001).

Upper threshold measures

We test for differences in prevalence reporting and exception reporting between practices that are above and below the upper threshold. For individual indicators we use a dummy variable U_t^{gki} (*upper threshold*) taking the value 1 if practice *g* had reported achievement (N_{gkit}/D_{gkit}) above the upper threshold α_{kiU} for indicator *i* in disease domain *k* in year *t.*

Prevalence reporting affects points and the value per point for all indicators in a domain. For models of prevalence reporting we include a measure of the average extent to which practice *g* was above the upper threshold in domain *k* in year *t*: $\bar{U}_t^{gk} = \sum_i \pi_{ki}^o U_t^{gki} / \sum_i \pi_{ki}^o$ (*domain upper*) where π_{ki}^o is the maximum number of points available for indicator *i* in disease domain *k*.

4.3 Estimation

Prevalence reporting

For 2005/6 prevalence reporting we estimate a pooled model for the indirectly age and sex standardised prevalence ratio

$$\left(100\frac{P_{gk}}{\hat{P}_{gk}}\right) = \beta_0 + \sum_j \beta_{1j}x_{gj} + \sum_\ell \beta_2 h_\ell + \gamma \bar{U}_{04/05}^{gk} + \varepsilon_{gk} \quad (2)$$

where $\hat{P}_{gk}$ is the expected number of prevalent patients for disease *k* in practice *g* given the age and sex composition of the practice. The Scottish NHS is administered through 15 Health Boards and we include Health Board dummies $h_{g\ell}$ to allow for differences in local policies or area characteristics not reflected in the covariates x_g. Since the dependent variable is an indirectly standardised ratio which has the same mean value (100) across disease domains we do not include disease domain dummies.

We estimate eight separate disease domain models

$$\left(100\frac{P_{gk}}{\hat{P}_{gk}}\right) = \beta_{0k} + \sum_j \beta_{1kj}x_{gj} + \sum_\ell \beta_{2k\ell} h_{g\ell} + \gamma_k \bar{U}_{04/05}^{gk} + \varepsilon_{gk} \quad (3)$$

To allow for the effects of the covariates and the domain upper threshold variable to differ across diseases.

[11] http://www.isdscotland.org/isd/servlet/controller?p_service=Content.show&p_applic=CCC&pContentID=1044

[12] $100*P_{gk}/\hat{P}_{gk} = 100*\sum_a p_{kag}\omega_{ag} / \sum_a p_{ka}^o \omega_{ga}$ where p_{ka}^o is the true prevalence rate for disease *k* in age and sex strata *a* in the reference sample practices, p_{kag} is the unobserved reported prevalence rate for disease *k* in age and sex strata *a* in practice *g*, ω_a is the proportion of sample practices' population in age and sex strata *a* and ω_{ag} is the proportion of the population of practice *g* in age and sex strata *a*.

We also estimate a model with practice fixed effects

$$\left(100\frac{P_{gk}}{\hat{P}_{gk}}\right)=\delta_0+\sum\nolimits_{g'}\delta_{1g'}D_{g'}+\gamma\bar{U}^{gk}_{04/05}+\varepsilon_{gk} \qquad (4)$$

where D_g is the dummy variable for practice g. The advantage of this specification is that the practice fixed effects will pick up all unobserved practice characteristics and thereby reduce the risk that the estimated coefficient on the domain upper threshold variables are biased. The disadvantage is that γ may be estimated imprecisely as it is based only on the variation in reported prevalence across the eight diseases domains with practices.

The three prevalence ratio specifications are estimated by Weighted Least Squares, using the practice populations as weights. Estimation of the standard errors for equation (11) allows for clustering within practices.

Exception reporting
We estimate a 2005/6 exception rate model pooled over the 65 indicators

$$\frac{E_{gki}}{P^T_{gk}}=\beta_0+\sum\nolimits_j\beta_{1j}x_{gj}+\sum\nolimits_\ell\beta_2h_{g\ell}+\sum\nolimits_{k'}\sum\nolimits_{i'}\beta_{3k'i'}d_{k'i'}+\gamma U^{gki}_{04/05}+\varepsilon_{gki} \qquad (5)$$

where P^T_{gk} = D_{gki} + E_{gki} is the sum of the number of patients declared suitable for the indicator and the number exception reported. We allow for the incentive effect of the upper threshold and the effects of the covariates to differ across the indicators by estimating separate models for the 65 indicators

$$\frac{E_{gki}}{P^T_{gk}}=\beta_{0ki}+\sum\nolimits_j\beta_{1kij}x_{gj}+\sum\nolimits_\ell\beta_{2ki}h_{g\ell}+\gamma_{ki}U^{gki}_{04/05}+\varepsilon_{gki} \qquad (6)$$

and we also estimate model with practice fixed effects

$$\frac{E_{gki}}{P^T_{gk}}=\delta_0+\sum\nolimits_{g'}\delta_{1g'}D_{g'}+\sum\nolimits_\ell\delta_{2ki}h_{g\ell}+\gamma U^{gki}_{04/05}+\varepsilon_{gki} \qquad (7)$$

We estimate these three exception reporting models using Weighted Least Squares, using P^T_{gk} = D_{gki} + E_{gki} as weights. However, there is a non-trivial proportion of zero exception rates and many rates are very small. We therefore also estimate a negative binomial model for E_{gki} using P^T_{gk} = D_{gki} + E_{gki} as the exposure term. We allow for clustering within practices in the pooled model (14).

5. Results

Although the Department of Health had forecast that practices would achieve 75% of the maximum points in the first year of the scheme, the mean score was over 90% and in 2005/6 it had increased to 97.5%, with 15% of practices achieving maximum points. Many practices treated more patients than necessary to achieve the maximum points on indicators. The percentage of cases where practices were above the upper threshold increased from 75.5% in 2004/5 to 91.8% in 2005/6. Practices above the upper threshold in 2005/6 could have reduced the number treated in 2005/6 by 11.8% without reducing their QOF revenue.[13]

[13] The averages are weighted by the maximum points achievable for each indicator. Thus the percentage of cases above the upper threshold in 2005/6 is $100\times\left(\sum_g\left(\sum_k\sum_i U^{gki}_{05/06}\pi^o_{ki}\Big/\sum_k\sum_i\pi^o_{ki}\right)\right)G^{-1}$ where $U^{gki}_{05/06}$ = 1 if practice g is above the upper threshold for indicator i in domain k, π^o_{ki} is the maximum points achievable for indicator i in domain k and G the total number of practices. The percentage reduction in numbers treated in cases above the threshold which would have been possible without reducing QOF revenue is

$$100\times\frac{\sum_k\sum_i\sum_g[n_{gki}]}{\sum_k\sum_i\sum_g N_{gki}}$$

where $[n_{gki}]$ is the integer part of $N_{gki}(\alpha_{gki}-\alpha_{Uki})/\ \alpha_{gki}$ where α_{gki} is achievement by practice g for indicator i in domain k, α_{kiU} is the upper threshold, N_{gki} is the number of patients for whom practice g achieved indicator i in domain k, and the sums are over practices above the threshold for the indicator

Table 4 shows, for the 65 clinical indicators, the maximum points available, the upper threshold, the percentage of practices achieving the upper threshold, practices' reported achievement (*N*/*D*), delivered quality (*N*/(*D+E*)) and exception reporting (*E*/(*D+E*). Mean practice exception reporting in 2005/6 ranges from 0.7% (BP3) to 27.6% (CHD10).

The percentage achieving the upper threshold in 2004/5 varied from 35.0% (DM13) to 98.6% (DM11). Delivered quality is lower than reported achievement, by virtue of its definition, but is also more variable across practices. Appendix D has results from regressions of reported achievement and delivered quality on practice and patient characteristics and shows that delivered quality was lower in practices with more income deprived patients, with a higher proportion from ethnic minorities, and with larger lists.

5.1 Reported prevalence

We hypothesised that practices that expect to be below the upper threshold may under report prevalence and those who expected to be above the threshold over report. Table 6 has results from two models of prevalence reporting for 2005/6, using data pooled across 8 disease domains. The coefficients on the covariates are similar across models and generally plausible. Reported prevalence is higher in practices with populations with a higher standardised illness ratio. Reported prevalence is higher in practices with younger GPs. It is lower in practices with larger populations and higher in practices with more GPs per patient, suggesting that such practices have a lower cost of case finding. SPICE participants and practices with two of the unsubsidised computer systems report higher prevalence.

The first model includes a variable *upper domain* ($\bar{U}^{gk}_{05/06}$) measuring the average extent to which the practice was above the upper threshold for the indicators in the disease domain in 2005/6. But whether the practice is above or below the upper threshold is endogenous, and so, as we argued in section 3.2, the coefficient on $\bar{U}^{gk}_{05/06}$ in a model of prevalence reporting for 2005/6 will be biased downward. The second model uses *lag upper domain* ($\bar{U}^{gk}_{04/05}$), measuring the extent to which the practice was above the upper thresholds in previous year. The coefficient on *lag upper domain* is positive and significant, suggesting that practices gamed prevalence reporting in 2005/6.

As we argued in section 3.2 the use of rich set of covariates makes it unlikely that the positive coefficient on $\bar{U}^{gk}_{04/05}$ is due to persistent unobservable factors increasing 2005/6 prevalence and increasing 2004/5 achievement. Moreover, such persistent factors should also increase 2005/6 achievement and produce a similar sized coefficients on $\bar{U}^{gk}_{05/06}$. The fact that the coefficient on $\bar{U}^{gk}_{05/06}$ in model 1 is negative whilst that on $\bar{U}^{gk}_{04/05}$ in model 2 is positive suggests that the negative coefficient on $\bar{U}^{gk}_{04/05}$ is evidence for gaming of exceptions, rather than arising from persistent unobservable factors.

The coefficient on $\bar{U}^{gk}_{04/05}$ in model 2 suggests that being above the upper threshold in 2004/5 increased reported standardised prevalence in 2005/6 by 3.5% of its mean.

Table 4. Ratio clinical indicators: rates of achievement, delivered quality, exception reporting. Scotland 2005/6

Indicator	Definition	Max points	Upper threshold (%)	% practices achieving threshold 2004/5		Reported upper achievement 2005/6 (%)		Delivered quality 2005/6 (%)		Exception reporting 2005/6 (%)		Effect on exception reporting rate 2005/6 of being above upper threshold in 2004/5	
												Negbin model	
		(1)	(2)	(3)		(4)		(5)		(6)		(7)	
				Mean	SD	Mean	SD	Mean	SD	Mean	SD	Coef	t
ASTHMA02	Diagnosis confirmed	15	70	80.2	39.8	88.8	9.7	83.2	11.9	6.4	7.2	-0.541	-5.880
ASTHMA03	Teenagers smoking record	6	70	79.5	40.4	85.3	10.3	80.1	11.2	6.1	8.0	-0.036	-0.302
ASTHMA04	Adults smoking record	6	70	98.3	13.1	94.9	4.1	92.4	4.5	2.6	3.0	-0.855	-2.008
ASTHMA05	Smoking cessation advice.	6	70	93.4	24.8	90.2	7.4	85.2	8.5	5.6	6.8	-0.298	-1.678
ASTHMA06	Reviewed in last 15 months	20	70	64.4	47.9	81.0	10.9	74.0	10.9	8.7	8.8	-0.087	-1.167
ASTHMA07	Received flu jab	12	70	51.1	50.0	78.3	11.1	57.5	8.5	26.6	11.3	0.074	2.572
BP02	Smoking record	10	90	90.3	29.6	98.3	1.7	97.1	2.2	1.2	1.2	-0.474	-5.037
BP03	Smoking cessation advice	10	90	83.6	37.1	98.3	2.4	97.6	2.8	0.7	1.4	-0.851	-6.181
BP04	BP recorded in last 9 months	20	90	59.8	49.1	93.4	4.5	91.4	4.6	2.2	2.4	-0.070	-1.049
BP05	BP ≤ 150/90	56	70	66.4	47.2	79.2	7.1	74.7	7.4	5.7	4.0	-0.256	-5.685
CANCER02	Reviewed within 6 months of diagnosis	6	90	65.9	47.4	93.7	8.9	87.3	10.1	6.8	5.6	-0.132	-2.202
CHD02	Exercise test or specialist assessment	7	90	57.1	49.5	91.5	12.0	84.2	13.3	8.0	7.8	-0.266	-3.553
CHD03	Smoking record	7	90	89.7	30.4	97.5	2.6	95.7	3.0	1.8	1.7	0.048	0.405
CHD04	Smoking cessation advice	4	90	74.9	43.4	95.8	4.6	92.7	5.4	3.3	3.5	-0.186	-2.160
CHD05	BP record in last 15 months	7	90	90.3	29.6	97.6	2.5	95.4	3.0	2.2	2.1	-0.144	-1.328
CHD06	BP ≤ 150/90	19	70	95.0	21.9	89.3	5.2	85.3	5.5	4.4	3.2	-0.264	-1.903
CHD07	Cholesterol measured in last 15 months	7	90	62.0	48.6	94.4	4.7	90.4	5.2	4.3	3.4	-0.010	-0.181
CHD08	Cholesterol ≤ 5mmol/l	16	60	80.5	39.7	80.5	8.6	72.0	7.8	10.6	5.5	0.003	0.061
CHD09	Takes Aspirin	7	90	58.9	49.2	95.4	3.2	92.2	3.5	3.4	2.6	-0.135	-2.609
CHD10	Treated with beta blocker	7	50	96.1	19.5	75.6	11.7	54.7	6.5	27.6	11.1	0.053	0.584
CHD11	Treated with ACE inhibitor	7	70	87.3	33.3	86.9	8.8	78.8	9.8	9.3	8.7	-0.299	-3.183
CHD12	Received flu jab	7	86	66.0	47.4	93.1	4.8	80.3	5.9	13.7	6.0	-0.056	-1.880
COPD02	Diagnosispost April 04 confirmed	5	90	55.2	49.8	91.9	11.6	84.4	13.8	8.2	8.0	-0.460	-7.041
COPD03	Diagnosis pre April 04 confirmed	5	90	48.1	50.0	92.9	12.1	82.5	13.4	11.2	9.3	-0.226	-3.751
COPD04	Smoking record	6	90	85.7	35.1	96.9	3.7	93.8	4.7	3.2	3.2	-0.158	-1.664
COPD05	Smoking cessation advice.	6	90	79.7	40.2	95.8	5.0	92.1	6.2	3.9	4.3	-0.106	-1.029
COPD06	FeV1 recorded in last 15 months	6	70	63.2	48.3	89.0	14.6	77.2	15.3	13.2	10.5	-0.260	-4.549
COPD07	Inhaled technique checked last 2 years.	6	90	49.1	50.0	92.2	10.2	85.3	11.2	7.5	6.5	-0.170	-2.736
COPD08	Flu jab	6	85	75.7	42.9	93.6	5.4	79.6	7.3	14.9	7.1	-0.107	-2.817
DM02	BMI record in last 15 months	3	90	82.6	38.0	96.0	3.2	92.5	4.0	3.6	2.8	-0.246	-3.589
DM03	Smoking recorded in last 15 months	3	90	97.1	16.9	98.5	1.8	97.0	2.4	1.5	1.5	-0.061	-0.250

DM04	Smoking cessation advice	5	90	79.2	40.6	96.1	4.9	92.8	6.2	3.4	3.9	-0.269	-2.510
DM05	HbA1c recorded in last 15 months	3	90	93.7	24.3	97.8	2.1	94.8	3.1	3.0	2.2	-0.265	-2.604
DM06	HbA1C ≤ 7.4	16	50	77.6	41.7	59.8	10.5	51.8	10.0	13.3	8.5	-0.166	-3.112
DM07	HbA1C ≤ 10	11	85	85.7	35.0	92.1	3.8	86.4	4.9	6.3	3.8	-0.135	-2.401
DM08	Retinal screening in last 15 months	5	90	51.8	50.0	92.9	7.1	86.1	7.7	7.2	5.4	-0.149	-3.287
DM09	Peripheral pulses recorded in last 15 months	3	90	47.1	49.9	90.7	7.5	84.8	8.6	6.5	4.9	-0.205	-4.513
DM10	Neuropathy test	3	90	43.2	49.6	90.0	7.9	84.0	9.0	6.6	5.0	-0.190	-4.219
DM11	BP record in last 15 months	3	90	98.6	11.8	98.7	1.4	96.9	2.2	1.8	1.7	-0.452	-2.143
DM12	BP ≤ 145/85	17	55	95.5	20.7	79.1	8.5	73.1	9.1	7.6	4.5	-0.467	-4.364
DM13	Micor-albuminuria record in last 15 months	3	90	35.0	47.7	87.5	11.0	81.4	11.6	7.0	5.9	-0.219	-4.035
DM14	Serum creatinine recorded in last 15 months	3	90	86.9	33.7	96.9	3.3	94.1	4.6	2.9	2.7	-0.231	-2.565
DM15	Treated with ACE inhibitor	3	70	82.2	38.3	85.2	10.1	78.1	11.0	8.3	8.6	-0.335	-3.338
DM16	Cholesterol record in last 15 months	3	90	86.7	34.0	96.7	2.6	93.9	3.4	2.9	2.3	-0.245	-3.437
DM17	Cholesterol < 5mmol/l	6	60	89.5	30.6	81.3	7.4	73.0	7.4	10.3	4.9	-0.078	-1.405
DM18	Flu jab	3	85	66.8	47.1	92.1	5.4	77.2	6.5	16.2	6.7	-0.020	-0.661
EPILEP02	Seizure frequency record in last 15 months	4	90	80.5	39.7	96.4	5.9	90.6	8.3	5.9	6.8	-0.105	-1.039
EPILEP03	Medication review in last 15 months	4	90	76.5	42.4	95.7	6.9	89.9	9.2	6.1	7.2	-0.119	-1.265
EPILEP04	Seizure free in last 12 months	6	70	36.5	48.2	71.2	14.1	55.8	13.0	21.6	14.4	0.103	2.073
LVD02	Diagnosis confirmed	6	90	62.9	48.4	94.1	10.3	88.2	12.5	6.3	8.3	-0.385	-3.569
LVD03	Treated with ACE	10	70	89.0	31.3	87.0	8.2	78.8	9.0	9.4	8.6	-0.477	-5.335
MH02	Reviewed in last 15 months	23	90	83.6	37.1	96.3	4.5	90.8	8.9	5.7	8.4	-0.222	-1.623
MH03	Lithium levels checked if on lithium	3	90	73.1	44.4	96.7	7.3	92.2	9.7	4.6	7.2	-0.194	-1.304
MH04	Seum creatinine /TSH checked if on lithium	3	90	75.4	43.1	95.9	9.5	91.9	11.7	4.1	6.6	-0.205	-1.316
MH05	Lithium level in therapeutic range	5	90	73.5	44.2	95.8	8.3	88.4	11.4	7.7	8.9	-0.137	-1.071
STROKE02	Confirmed by CT / MRI	2	80	78.0	41.4	92.2	9.7	85.8	11.7	7.0	7.2	-0.417	-4.431
STROKE03	Smoking record in last 15 months	3	90	84.1	36.6	96.4	3.4	93.4	4.3	3.1	3.2	-0.191	-2.077
STROKE04	Smoking cessation advice	2	70	94.4	23.1	93.5	6.3	89.4	7.5	4.3	5.3	-0.326	-1.486
STROKE05	BP record in last 15 months	2	90	85.7	35.1	97.0	2.9	93.7	4.1	3.4	3.5	-0.264	-2.766
STROKE06	BP < 150/90	5	70	90.0	30.0	87.9	6.0	81.9	6.8	6.8	4.9	-0.377	-4.541
STROKE07	Cholesterol record in last 15 months	2	90	51.5	50.0	92.6	6.2	86.1	7.2	7.0	5.5	-0.023	-0.411
STROKE08	Chol. < 5mmol/l	5	60	65.0	47.7	76.8	9.6	65.5	9.3	14.7	7.8	-0.071	-1.716
STROKE09	Aspirin treatment	4	90	61.3	48.7	94.9	3.8	90.7	4.7	4.4	3.6	-0.172	-2.880
STROKE10	Flu jab	2	85	56.0	49.7	90.5	6.3	75.5	7.3	16.6	7.4	-0.019	-0.639
THYROI02	Thyroid function test in last 15 months	6	90	91.3	28.2	96.9	3.0	96.0	3.1	0.9	1.2	-0.102	-0.664

Reported achievement: *N*/*D*; delivered quality: *N*/(*D+E*); Exception reporting: *E*/(*D+E*). Denominator weighted means over Scottish practices.
Separate Negbin models for each indicator also contain all variables in pooled exception rate model except for indicator dummies.

Table 5. Summary statistics: explanatory variables

Variable	Definition	Mean	SD	Min	Max
Patient characteristics					
Prop ≤15	Proportion of patients aged 15 or less	0.165	0.028	0.026	0.270
Prop >75	Proportion of patients over 75 years	0.069	0.020	0.007	0.159
SIR	Standardised illness ratio	98.842	22.619	50.533	186.742
SMR	Standardised mortality ratio	107.991	25.092	44.199	329.250
Income depriv	IMD income deprivation score	15.050	7.519	2.801	43.409
Education depriv	IMD education deprivation score	0.001	0.591	-1.583	1.671
Employment depriv	IMD employment deprivation score	14.038	6.305	3.089	36.466
Housing depriv	IMD housing deprivation score	19.855	10.981	4.900	60.313
Access depriv	IMD access deprivation score	-0.020	0.446	-1.118	2.719
Ethnicity	Minority ethnic group proportion	0.021	0.028	0.000	0.408
Rural	Population mode in settlements <3,000 people	0.101	0.301	0.000	1.000
Sparsity	Inverse of population density (hectares per person)	1.354	4.625	0.014	117.289
Practice characteristics					
List 1000	List size in 000s	5.471	3.251	0.174	19.171
GP/list 1000	GPs per 1,000 patients	0.973	0.710	0.263	8.000
Prop female GP	Proportion of female GPs	0.423	0.252	0.000	1.000
Av GP age	Average age of GPs (years)	44.747	5.801	28.250	67.000
Prop GP not UK qual	Proportion of GPs not qualified in UK	0.092	0.217	0.000	1.000
PMS contract	PMS contract	0.092	0.290	0.000	1.000
Dispensing	Practice dispenses medicines	0.108	0.311	0.000	1.000
Training	Practice is a training practice	0.285	0.452	0.000	1.000
Ex FH	Ex fundholding practice	0.461	0.499	0.000	1.000
New practice	Practice formed since 1999	0.065	0.247	0.000	1.000
Non-FH	Practice was not a fundholder (baseline)	0.474	0.500	0.000	1.000
Herfindahl	Herfindahl index for areas served	0.353	0.244	0.083	0.998
SPICE	Practice in SPICE network in Autumn 2004	0.329	0.470	0.000	1.000
Patient Inflow	Proportion patients registered in last year	0.069	0.042	0.006	0.845
Software: Egton	Uses Egton Medical Information Systems software	0.054			
Software: GPASS	Uses GPASS software (baseline)	0.863			
Software: In Practice	Uses In Practice Systems software	0.053			
Software: Protechnic	Uses Protechnic Exeter software	0.009			
Software: iSoft	Uses iSoft software	0.022			

Based on 916 practices. Patient characteristics are weighted by list size. Practice characteristics are unweighted.

Table 6 Determinants of prevalence reporting in Scottish practices 2005/6: pooled models

Dependent variable	Standardised reported prevalence ratio 2005/6[1]			
	Coef.	t	Coef.	t
dom upper	**-3.299**	**-1.51**		
dom upper_lag			**3.461**	**3.07**
SIR	0.501	4.08	0.531	4.32
SMR	0.010	0.42	0.011	0.43
Income depriv	0.095	0.35	0.114	0.42
Education depriv	10.255	5.00	10.027	4.90
Employment depriv	-0.492	-1.10	-0.548	-1.24
Housing depriv	-0.175	-1.97	-0.190	-2.18
Access depriv	2.438	1.69	2.082	1.46
Ethnicity	-52.386	-1.63	-52.220	-1.63
Rural	-1.131	-0.72	-1.283	-0.83
Sparsity	-0.366	-4.87	-0.367	-4.96
List 1000	-0.475	-2.58	-0.484	-2.68
GP/list 1000	5.163	3.82	5.246	3.91
Prop female GP	4.038	2.14	3.764	2.01
Av GP age	-0.313	-3.67	-0.298	-3.51
Prop GP not UK qual	0.796	0.31	0.936	0.37
PMS contract	0.957	0.71	1.138	0.85
Dispensing	0.170	0.10	0.302	0.19
Training	0.391	0.46	0.314	0.37
Ex FH	0.792	0.98	0.581	0.72
New practice	2.046	1.40	1.911	1.32
Herfindahl	2.884	1.28	3.474	1.57
SPICE	3.345	4.23	3.100	3.95
Patient Inflow	-19.999	-1.64	-18.885	-1.58
Software: Egton	2.885	1.95	2.530	1.70
Software: In Practice	4.941	3.24	4.464	2.98
Software: Protechnic	-1.392	-0.53	-1.506	-0.59
Software: iSoft	2.629	0.89	2.322	0.80
Constant	70.498	7.22	61.603	6.62
$\bar{R}^2$	0.2407		0.2416	
N	7449		7449	

[1] Age and sex indirectly standardised reported prevalence ratio 2005/6.
N is number of number of disease domain - practice observations
t stats are robust and allow for clustering within practices. OLS models are weighted by population size and include 14 NHS Board dummies.
Domain upper: proportion of indicators in domain for which practice achieved upper threshold in 2005/6 weighted by max points for indicators
Lag domain upper: proportion of indicators in domain for which practice achieved upper threshold in 2004/5 weighted by max points for indicators.

Pooled models with alternative sets of covariates (not shown) produced similar coefficients on *lag upper domain.* However, the pooled model (2) forces the effects of the covariates to be the same for all disease domains. Given the differences in aetiology this is problematic. We therefore report in Table 7 the coefficients on *lag upper domain* from separate models (3) for standardised prevalence reporting in 2005/6 for the eight disease domains. The effect of *lag upper domain* is positive in seven of the eight domains and is significant in five of them.

Table 7. Determinants of prevalence reporting in Scottish practices 2005/6

Disease:	CHD	Diabetes	Asthma	COPD	Epilepsy	BP	Thyroid	Stroke
Lag domain upper	1.721	7.917	7.782	-0.4	2.079	2.938	5.635	6.876
(t stat)	0.718	3.499	3.178	-0.107	0.869	1.973	2.096	2.262
$\overline{R}^2$	0.6243	0.6145	0.1793	0.6230	0.4397	0.2381	0.5354	0.3538
N	932	932	932	932	925	932	932	932

Dependent variables: age and sex indirectly standardised disease reported prevalence ratio 2005/6. All models included the covariates in Table 7 plus 14 NHS Board dummies. OLS coefficient reported above robust t statistic adjusted for clustering within practices.
Lag domain upper: proportion of indicators in domain for which practice achieved upper threshold in 2004/5 weighted by max points for indicators.

We also estimated the pooled 2005/6 prevalence reporting model (4) with practice fixed effects. The coefficient on *lag upper domain* is again positive 1.627 (t =1.20) but smaller than in Table 6 and not significant. By including practice fixed effects in the regression model we remove all unobservable practice variables but leave only the variation of prevalence across 8 domains within the practice to be explained by the *lag upper domain* variable. Since practices which were above indicator thresholds in one disease domain tended to be above them in other disease domains, there is relatively little variation in *lag upper domain* within practices. It is therefore perhaps not surprising that the coefficient on lag upper domain is not precisely estimated.

5.2 Exception reporting

Table 8 has the results for the OLS and NegBin regressions of 2005/6 exception reporting pooled over 65 indicators. The coefficients on the patient characteristic covariates are similar in the two models and are plausible. Exception reporting is lower in practices with older populations, higher in practices with sicker patients and with populations drawn from areas with a higher ethnic minority proportion. Exception rates are lower in rural areas.

Since the rate of true exceptions are is determined by patient characteristics, our first test for gaming is whether the reported exception rate varies with practice characteristics which ought not to affect true exceptions. Table 8 shows that practices that were fundholding have higher exception reporting. Training practices have lower exception reporting. SPICE participants and practices with higher Herfindahls have higher exception reporting. These results suggest that practices were gaming their exceptions.

Our more direct test of gaming uses the difference in incentives for practices above and below the upper threshold. The coefficients on the *upper threshold* ($U_{05/06}^{gki}$) dummy (which equals 1 if the practice is above the upper threshold for the indicator in 2005/6) are positive and highly significant. We interpret this as evidence of endogeneity bias since errors unobserved factors increasing 2005/6 exceptions will also increase achievement and hence make it more likely that the practice is above the upper threshold in 2005/6.

The models using *Lag upper threshold* ($U_{04/05}^{gki}$) provide strong evidence of gaming: the coefficients on *Lag upper threshold* are negative and highly significant. Pooled models with subsets of the deprivation measures and including the measure of self assessed health produced similar coefficients on *lag upper threshold*.

For the 14,384 of the 56,980 practice-indicator cases where achievement was below the threshold in 2004/5 the mean exception rate in 2005/6 was 8.55%. Using the NegBin results for model 2 we calculate that if achievement had been above the upper threshold in 2004/5 reported exceptions would have been 7.25%. Thus the incentive to overstate exceptions lead to non-trivial gaming in the 25% of practice-indicator cases where practices were below the upper threshold in 2004/5.

Table 8. Determinants of exception reporting in Scottish practices 2005/6: pooled regressions

	OLS models[1]				Negative binomial models[2]			
	Coef	t	Coef	t	Coef	t	Coef	t
Upper threshold	1.025	6.47			0.092	4.88		
Lag upper threshold			-0.729	-6.91			-0.165	-11.75
Prop ≤15	2.609	0.89	3.541	1.2	-0.084	-0.2	0.095	0.22
Prop >75	-8.588	-2.66	-9.326	-2.88	-1.117	-2.22	-1.154	-2.29
SIR	0.005	0.41	-0.003	-0.28	-0.002	-1.29	-0.003	-1.9
SMR	0.004	1.08	0.004	1.18	0.001	1.89	0.001	2.03
Income depriv	0.103	2.6	0.094	2.39	0.024	3.82	0.021	3.43
Education depriv	-0.757	-2.74	-0.731	-2.66	-0.108	-2.63	-0.106	-2.59
Employment depriv	-0.069	-1.19	-0.046	-0.81	-0.014	-1.62	-0.010	-1.15
Housing depriv	0.032	2.52	0.038	2.97	0.002	0.84	0.003	1.34
Access depriv	-0.223	-1.04	-0.194	-0.91	-0.069	-2.14	-0.059	-1.84
Ethnicity	9.912	2.68	9.214	2.57	0.965	2.22	0.920	2.14
Rural	-0.936	-4.8	-0.908	-4.68	-0.167	-4.94	-0.161	-4.77
Sparsity	-0.018	-1.83	-0.018	-1.79	-0.002	-1.6	-0.002	-1.51
List 1000	-0.007	-0.37	-0.008	-0.41	0.004	1.26	0.004	1.32
GP/list 1000	0.235	1.35	0.246	1.41	0.070	2.88	0.068	2.78
Prop female GP	-0.152	-0.58	-0.080	-0.31	0.014	0.39	0.028	0.76
Av GP age	0.008	0.66	0.005	0.43	0.000	0.06	0.000	-0.28
Prop GP not UK qual	0.193	0.56	0.134	0.39	0.055	1.12	0.056	1.13
PMS contract	0.510	2.85	0.482	2.68	0.045	1.7	0.037	1.39
Dispensing	-0.311	-1.26	-0.345	-1.4	-0.055	-1.19	-0.065	-1.43
Training	-0.204	-1.52	-0.200	-1.49	-0.046	-2.16	-0.040	-1.92
Ex FH	0.492	4.15	0.540	4.55	0.066	3.7	0.073	4.06
New practice	0.139	0.68	0.169	0.82	0.007	0.22	0.011	0.35
Herfindahl	0.947	2.92	0.847	2.63	0.127	2.43	0.109	2.11
SPICE	0.272	2.18	0.328	2.63	0.049	2.65	0.059	3.21
Patient Inflow	-0.756	-0.63	-1.083	-0.89	-0.062	-0.28	-0.081	-0.37
Software: Egton	-0.642	-3.06	-0.582	-2.78	-0.004	-0.09	0.007	0.18
Software: In Practice	-0.747	-3.2	-0.643	-2.8	-0.114	-3.42	-0.093	-2.85
Software: Protechnic	0.709	1.09	0.702	1.09	0.170	1.83	0.160	1.75
Software: iSoft	-0.542	-2.03	-0.465	-1.77	-0.037	-0.77	-0.023	-0.5
Constant	2.322	1.86	4.531	3.63	-3.051	-16.89	-2.775	-15.49
$\bar{R}^2$	0.6011		0.6012					
Initial Log L					-166132.07		-166132.07	
Model Log L					-147920.48		-147762.1	
N	56980		56980		56980		56980	

[1] OLS models: dependent variable is the exception reporting rate [100*E/(D+E)]
[2] Negative binomial regressions of E using (D+E) as the exposure term
All models also contain 14 NHS Board dummies and 64 indicator dummies. Robust t statistics allow for clustering of indicators in practices.
Upper threshold = 1 if practice above upper threshold in 2005/6 for indicator, 0 otherwise.
Lag upper threshold = 1 if practice above upper threshold in 2004/5 for indicator, 0 otherwise

The fact that the coefficient on $U_{04/05}^{gki}$ in model 2 is negative and significant whereas the coefficient on $U_{05/06}^{gki}$ in model 1 is positive and significant suggests that the negative coefficient on $U_{04/05}^{gki}$ is due to gaming rather than to persistent unobservable factors negatively correlated with 2004/5 achievement and 2005/6 exceptions.

We also estimated separate OLS and NegBin models (6) for the 65 indicators. We report the NegBin coefficients on *Lag upper threshold* from these models in column (7) of Table 4. Only five of the indicators have positive coefficients on *Lag upper threshold* and two of these is significant. All the other 60 coefficients are negative and 36 of them are significant.

Models with practice fixed effects also provide strong evidence of gaming. The coefficient on *Lag upper threshold* in the NegBin model was –0.081 with a *z* statistic of 12.05.

6 Conclusions

The introduction of the QOF may have had its intended consequence: the first year of the QOF seems to show above trend performance against some clinical indicators. Practice behaviour is compatible with quasi-altruism: overall practices could have reduced the number of patients treated by 11.8% without reducing their revenue from the QOF.

Delivered quality was inequitable with respect to the income and ethnicity of the populations in the areas from which practices drew their lists. This is in contrast with consultations with general practitioners: allowing for morbidity, income has no effect on consultations, and some ethnic minority groups have more than expected numbers of consultations (Morris, Sutton and Gravelle, 2005). Similar results were obtained in earlier studies of the first year of the QOF in Scotland (McLean et al, 2006) and England (Doran et al, 2006). However in the absence of pre-QOF evidence on the relationship between the QOF quality indicators and socio-economic characteristics it is unclear whether the introduction of the QOF increased or reduced inquity. Practices with more female GPs and with younger GPs deliver higher quality.

There is also evidence of gaming. First, both true prevalence and reported exceptions are determined by patient characteristics. But we found that they also varied with practice characteristics such as the average age of the GPs in the practice or the total number of patients, or whether the practice had held a budget under the fundholding scheme. These are factors which plausibly affect GP costs or reflect their preferences and hence their reporting decisions but which should not affect true prevalence or exceptions.

Second, differences in the behaviour of practices above and below the upper threshold for indicators, where there is a sharp discontinuity in reporting incentives, also suggests gaming of reporting. Practices which were above the upper threshold in 2004/5 had reported standardised prevalence in 2005/6 which was 3.5% greater, against a mean of 100, than if they had been below the upper threshold in 2004/5. Practices below the upper threshold in 2004/5 had an average exception rate of 8.55% in 2005/6 and we estimate that without the incentive to increase their exceptions they would have had an exception rate of 7.25 %.

The ratio performance indicators in the QOF were intended to incentivise doctors to increase the numerator by treating more patients. But ratio indicators also create incentives to manipulate the denominators. Earlier more limited quality incentive schemes paid practices fixed sums only if more than a specified percentage of women aged 25 to 64 were screened for cervical cancer or if more than they vaccinated more than a specified percentage of children on their list. The earlier schemes did not permit exception reporting so that the only way GPs could reduce the denominator to achieve the target was to remove patients from their lists. GPs recognised the financial incentive to manipulate the denominator (Pickin et al, 2001) but, perhaps because of the more severe implications for the removed patients, there were few documented cases of them doing so. The QOF makes it easier for practices to manipulate the denominator without direct harm to patients and our evidence suggests that some practices did so.

References

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

Custers T, Arah O, Klazinga N. Is there a business case for quality in The Netherlands? A critical analysis of the recent reforms of the health care system. *Health Policy* 2006. In press, available online: doi:10.1016/j.healthpol.2006.09.005.

Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. "Family practice performance in the first year of the UK's new 'pay for performance' scheme: good clinical practice or gaming?" *New England Journal of Medicine*, 2006, 355, 375-384.

Dusheiko M, Gravelle H, Jacobs R, Smith PC. 2006. The effect of budgets on doctor behaviour: evidence from a natural experiment. *Journal of Health Economics* 25, 449-478.

Epstein A. 2006. Paying for performance in the United States and abroad. *New England Journal of Medicine*. 355, 406-408.

Fisher ES. 2006. Paying for performance – risks and recommendations. *New England Journal of Medicine*. 355, 1845-1847

Gené-Badia J, Escaramis-Babiano G, Sans-Corrales M, Sampietro- Colom L, Aguado-Menguy F, Cabezas-Peña C, Gallo-Puelles P. Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care. *Health Policy*, 2007;80:2-10

Greb S. Focke A, Hessel F, Wasem J. Financial incentives for disease management programmes and integrated care in German social health insurance. *Health Policy*, 2006; 78(2-3): 295-305.

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

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

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, version 1.1, R22 HSCIC. http://www.ic.nhs.uk/statistics-and-data-collections/audits-and-performance /the-quality-and-outcomes-framework-qof

Holmstrom B, Milgrom P. 1991. Multi-task principle-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7, 24-52.

Institute of Medicine. 2007. *Rewarding Provider Performance: Aligning Incentives in Medicare*. The National Academies Press. http://books.nap.edu/openbook.php?isbn=0309102162&page=R1. Accessed 27 July 2007.

McLean G, Sutton M, Guthrie B, 2006. Deprivation and quality of primary care services. Evidence for persistence of the inverse care law from the UK quality and outcomes framework. *Journal of Epidemiology and Community Health* 60, 917 - 922.

McGuire T. 2000. Physician agency. In Culyer, A. and Newhouse, J. (eds.). *Handbook of Health Economics*, North Holland, Amsterdam.

Medicare Australia. Practice Incentives Program. Available at http://www.medicareaustralia.gov.au/providers/incentives_allowances/pip.shtml Accessed 17 July 2007.

Morris S, Sutton, M, Gravelle H, 2005. Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science and Medicine* 60, 1251-1266.
Perkins R, Seddon M, on behalf of EPIQ. 2006. Quality improvement in New Zealand healthcare. Part 5: measurement for monitoring and controlling performance--the quest for external accountability. *New Zealand Medical Journal*. 119: U2149.

Pickin M, Sampson F, Munro J, Nichol J. 2001. General practitioners' reasons for removing patients from their lists: postal survey in England and Wales. *British Medical Journal*. 322, 1158-1159.

Pink G, Brown A, Studer M, Reiter K, Leatt P. 2006. Pay for performance in publicly financed healthcare: some international experience and considerations for Canada. *Healthcare Papers*. 6, 9-26.

Prendergarst C. 1999. The provision of incentives in firms. *Journal of Economic Literature*. 37(1), 7-63.

Roland M, 2004. Linking physicians' pay to the quality of care – a major experiment in the United Kingdom. *New England Journal of Medicine* 351, 1448-1454.

Rosenthal M, Frank R, Li Z, Epstein A. 2005. Early experience with pay-for-performance: from concept to practice. *Journal of the American Medical Association*, 294, 1788-93

Rosenthal M, Landon B, Normand S, Frank R, Epstein M. 2006. Pay for performance in commercial HMOs. *New England Journal of Medicine,* 355, 1895-902

Scott IA. 2007. Pay for performance in health care: strategic issues for Australian experiments. *Medical Journal of Australia*, 187(1), 2 July, 31-35.

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.

Whalley D, Gravelle H, Sibbald B. Impact of the new general medical services contract on general practitioners' job satisfaction and perceptions of quality of care in the UK. *British Journal of General Practice*, in press.

Appendices

A. Grounds for exception reporting

A) patients who have been recorded as refusing to attend review who have been invited on at least three occasions during the preceding twelve months
B) patients for whom it is not appropriate to review the chronic disease parameters due to particular circumstances eg terminal illness, extreme frailty
C) patients newly diagnosed within the practice or who have recently registered with the practice, who should have measurements made within three months and delivery of clinical standards within nine months eg blood pressure or cholesterol measurements within target levels
D) patients who are on maximum tolerated doses of medication whose levels remain sub-optimal
E) patients for whom prescribing a medication is not clinically appropriate eg those who have an allergy, another contraindication or have experienced an adverse reaction
F) where a patient has not tolerated medication
G) where a patient does not agree to investigation or treatment (informed dissent), and this has been recorded in their medical records
H) where the patient has a supervening condition which makes treatment of their condition inappropriate eg cholesterol reduction where the patient has liver disease
I) where an investigative service or secondary care service is unavailable.

Source: GMS Statement of Financial Entitlements, 30 March 2005.

B. Reported prevalence and revenue

Notation

P_{gk}	prevalent patients reported in disease domain k by practice g
D_{gki}	patients declared suitable for indicator i in domain k
$E_{gki} = P_{gk} - D_{gki}$	number of patients exception reported for indicator i
N_{gki}	number of patients for whom indicator i achieved
$\alpha_{gki} = N_{gki}/D_{gki}$	reported achievement rate for indicator i
α_{kiL}	lower achievement threshold for indicator i
α_{kiU}	upper achievement threshold for indicator i
$\pi_{gki}=\pi_{ki}(\alpha_{gki})$	points achieved for indicator i
π^{o}_{ki}	maximum points achievable for indicator i
M_g	practice population
V	national average price (value) per point
$v_{gk} = vF_{gk} M_g / \bar{M}$	value per point for indicators in domain k in practice g
$F_{gk} = \left(\frac{P_{gk}}{M_g}\right)^{\frac{1}{2}} \div \left(\frac{1}{G}\sum_h^G \left(\frac{P_{hk}}{M_h}\right)^{\frac{1}{2}}\right)$	adjusted disease prevalence factor for domain k for practice g
$R_{gki} = \pi_{gki} v_{gk}$	revenue from indicator i in domain k.

Practice g revenue from indicator i in disease domain k is

$$R_{gki} = v\left[M_g / \bar{M}\right]\pi_{ki}(\alpha_{gki})\left[P_{gk}/M_g\right]^{\frac{1}{2}}\left[G^{-1}\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1}$$
$$= \bar{v}_g \pi_{ki}(\alpha_{gki}) F_{gk}(P_{gk}) \qquad (1)$$

where $\pi_{gki} = \pi_{ki}(\alpha_{gki}) = \pi^{o}_{ki}\min\{1, \max\{(\alpha_{gki} - \alpha_{kiL})/(\alpha_{kiU} - \alpha_{kiL}), 0\}\}$ and $\alpha_{gki} = N_{gki}/(P_{gk} - E_{gki})$.

For $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$, $\partial R_{gki}/\partial P_{gk}$ has the same sign as

$$-\frac{\alpha_{gki}}{(P_{gk}-E_{gki})}P_{gk}^{\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1}$$
$$+ (\alpha_{gki}-\alpha_{kiL})\left\{\frac{1}{2}P_{gk}^{-\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1} - P_{gk}^{\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-2}\frac{1}{2}P_{gk}^{-\frac{1}{2}}M_g^{-\frac{1}{2}}\right\}$$
$$< -\frac{\alpha_{gki}}{(P_{gk}-E_{gki})}P_{gk}^{\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1} + (\alpha_{gki}-\alpha_{kiL})\frac{1}{2}P_{gk}^{-\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1}$$
$$= \left[-\frac{\alpha_{gki}}{(P_{gk}-E_{gki})} + (\alpha_{gki}-\alpha_{kiL})\frac{1}{2P_{gk}}\right]P_{gk}^{\frac{1}{2}}\left[\sum_h (P_{hk}/M_h)\right]^{-1} < 0 \qquad (2)$$

C. Comparative statics

The practice objective function is

$$s(N,E,P) = bN + R(N,E,P) - C(N,E,P)$$

$$= bN + \bar{v}\pi(\alpha)F(P) \;\; - c_{N1}N - c_{E1}\left(\left|E - E^o\right| / P^o\right) - c_{P1}\left(\left|P - P^o\right| / M\right)$$

$$- \tfrac{1}{2}\left[c_{N2}N^2 + c_{E2}\left((E - E^o)/P^o\right)^2 + c_{P2}\left((P - P^o)/M\right)^2\right] \tag{1}$$

R is twice differentiable except at α_L where the derivative of points $\pi(\alpha)$ with respect to achievement π' jumps from zero to positive and at α_U where it jumps down to zero. C is also twice differentiable except at E^o (and P^o) where the derivative with respect to E (and P) jumps from negative to positive.

Proposition 1. For any given vector of cost parameters, true prevalence, true exceptions and list size, there exist b_1, b_2, b_3, b_4 (with $b_1 < b_2 < b_3 < b_4$) such that for all $b < b_1$ the practice optimum has $\alpha < \alpha_L$, for $b \in (b_2, b_3)$ the optimum has $\alpha \in (\alpha_L, \alpha_U)$ and for $b > b_4$ the optimum has $\alpha > \alpha_U$.

Solutions with $\alpha < \alpha_L$

The first order conditions are

$$s_N = b - C_N \le 0, \quad N \ge 0, \quad s_N N = 0 \tag{2}$$

$$s_E^+ = -C_E^+ > 0 > s_E^- = -C_E^- \tag{3}$$

$$s_P^+ = -C_P^+ > 0 > -C_P^- \tag{4}$$

where superscripts +, – indicate left and right sided derivatives with respect to E evaluated at $E^{*L} = E^o$ and similarly for P at $P^{*L} = P^o$. The solution has $E^{*L} = E^o$, $P^{*L} = P^o$ since gaming generates no additional revenue when $\alpha < \alpha_U$. Since $s_N(0,E^o,P^o) = b - C_N(0,E^o,P^o) = b - c_{P1}$ and c_{P1} is finite there always exists a sufficiently high b (b_1) (or small c_{P1}) such that $N^{*L}(b) > 0$ for $b > b_1$. It is also obvious that $N^{*L}(b)$ (and hence $\alpha^{*L} = N^{*L}/(P^{*L} - E^{*L})$) is increasing in b.

Solutions with $\alpha \in (\alpha_L, \alpha_U)$

Consider first solutions with $\alpha \in (\alpha_L, \alpha_U)$ and where the cost parameters c_{E1} and c_{P1} are small enough that the optimal $E^{*UL} \neq E^o$ and $P^{*UL} \neq P^o$. The first order conditions are

$$s_N = b + R_N - C_N = b + \bar{v}\pi'\alpha_N F - C_N = 0 \tag{5}$$

$$s_E = R_E - C_E = \bar{v}\pi'\alpha_E F - C_E = 0 \tag{6}$$

$$s_P = R_P - C_P = \bar{v}\pi'\alpha_P F + \bar{v}\pi(\alpha)F' - C_P = 0 \tag{7}$$

Recall that when $\alpha \in (\alpha_L, \alpha_U)$, $R_E > 0$ and $R_P < 0$, so that the solution has $C_E > 0$ and $C_P < 0$ which requires $E^{*UL} > E^o$ and $P^{*UL} < P^o$.

The second order partials of $s(N,E,P)$ are

$$s_{NN} = -C_{NN} < 0 \tag{8}$$

$$s_{EE} = \bar{v}\pi'\alpha_{EE}F - C_{EE} < 0 \tag{9}$$

$$s_{PP} = \bar{v}\pi'\alpha_{PP}F + 2\bar{v}\pi'\alpha_P F' + \bar{v}\pi F'' - C_{PP} < 0 \tag{10}$$

$$s_{NE} = \bar{v}\pi'\alpha_{NE}F > 0 \tag{11}$$

$$s_{NP} = \bar{v}\pi'\alpha_{NP}F + \bar{v}\pi'\alpha_N F'$$

$$= N^{-1}\left[\bar{v}\frac{\pi^o}{\alpha_U - \alpha_L}\alpha_P F + \bar{v}\frac{\pi^o(\alpha - \alpha_L)}{\alpha_U - \alpha_L}F' + \bar{v}\frac{\pi^o\alpha_L}{\alpha_U - \alpha_L}F'\right]$$

$$= N^{-1}\left[R_P + \bar{v}\frac{\pi^o\alpha_L}{\alpha_U - \alpha_L}F'\right] < 0 \tag{12}$$

$$s_{EP} = \bar{v}\pi'\alpha_{EP}F + \bar{v}\pi'\alpha_E F' = \bar{v}\pi' 2\alpha_{NP}\alpha F + \bar{v}\pi'\alpha_N \alpha F'$$
$$= \alpha\left[s_{NP} + \bar{v}\pi' 2\alpha_{NP}\alpha F\right] < 0 \qquad (13)$$

The sign of s_{NN} follows from the convexity of the cost function. Since $\alpha_{EE} = \alpha_{PP} > 0$, the signs of s_{EE} and s_{PP} require the stronger restrictions that marginal costs of E and P are increasing faster than the marginal revenue from E and P. $s_{NE} > 0$ follows from $\alpha_{NE} > 0$. $s_{NP} < 0$ since adding the term $\bar{v}\pi^o\alpha_L(\alpha_U - \alpha_L)^{-1}F'$ to R_p changes the square bracketed term in (2) to $-\frac{\alpha_{gki}}{(P_{gk} - E_{gki})} + \frac{\alpha_{gki}}{2P_{gk}} < 0$. Finally $s_{EP} < 0$ because $s_{NP} < 0$ and $\alpha_{NP} < 0$.

The signs of the cross partials are not sufficient to ensure that $s(N,E,P)$ is strictly concave but they do ensure some definite comparative static results. Since $s_{Nb} = 1$, $s_{Eb} = 0$, $s_{Pb} = 0$ we have

$$\frac{\partial N^{*UL}}{\partial b} = -\left[s_{EE}s_{PP} - (s_{PE})^2\right]\Delta^{-1} > 0 \qquad (14)$$

$$\frac{\partial E^{*UL}}{\partial b} = \left[s_{EN}s_{PP} - s_{PN}s_{EP}\right]\Delta^{-1} > 0 \qquad (15)$$

$$\frac{\partial P^{*UL}}{\partial b} = -\left[s_{EN}s_{PE} - s_{PN}s_{EE}\right]\Delta^{-1} < 0 \qquad (16)$$

where the Hessian $\Delta < 0$ since $s(N,E,P)$ is locally concave in the neighbourhood of the solution. Local concavity also implies that the principal minor $s_{EE}s_{PP} - (s_{PE})^2 > 0$, hence establishing the sign of (14). The signs (15) and (16) follow from $\Delta < 0$ and the signs of the cross partials previously established. Since $\partial N^{*UL}/\partial b > 0$ and $\frac{\partial P^{*UL}}{\partial b} - \frac{\partial E^{*UL}}{\partial b} < 0$ we also have $\partial\alpha^{*UL}/\partial b > 0$.

Given the separability of cost function in N, E, and P similar arguments establish that an increase in c_{N1} or c_{N2} reduces N^{*UL} and E^{*UL}, increases P^{*UL}, and hence reduces α^{*UL}.

Now consider solutions with $\alpha \in (\alpha_L, \alpha_U)$ and the cost parameter c_{E1} large enough that the optimal $E^{*UL} = E^o$ so that the first order conditions are

$$s_N = b + R_N - C_N = b + \bar{v}\pi'\alpha_N F - C_N = 0 \qquad (17)$$

$$s_E^+ = R_E - C_E^+ > 0 > s_E^- = R_E - C_E^- \qquad (18)$$

$$s_P = R_P - C_P = \bar{v}\pi'\alpha_P F + \bar{v}\pi(\alpha)F' - C_P = 0 \qquad (19)$$

where s_E^+, s_E^- are the left and right sided derivatives of s with respect to E evaluated at E^o. Then we can use the local concavity of s in N and P and the signs of s_{NN} and s_{NP} to establish that $\partial N^{*UL}/\partial b > 0$, $\partial P^{*UL}/\partial b < 0$ and $\partial\alpha^{*UL}/\partial b > 0$. If the first order condition (18) on E is replaced by

$$s_E^+ = R_E - C_E^+ > 0 = s_E^- = R_E - C_E^- \qquad (20)$$

Then an increase in b will also increase E^{*UL} from E^o. (Suppose not. But then the increase in α^{*UL} via the increase in N^{*UL} and the reduction in P^{*UL} will increase $R_E = \bar{v}\pi^o(\alpha_U - \alpha_L)^{-1}N(P - E^o)^{-2}$ and $s_E^- > 0$.)

Similar arguments apply to cases where $\alpha \in (\alpha_L, \alpha_U)$ with $P^{*UL} = P^o$ and where $\alpha \in (\alpha_L, \alpha_U)$ with $E^{*UL} = E^o$, $P^{*UL} = P^o$. Thus for all solutions with $\alpha \in (\alpha_L, \alpha_U)$ we have established that $\partial N^{*UL}/\partial b > 0$, $\partial E^{*UL}/\partial b \geq 0$, $\partial P^{*UL}/\partial b \leq 0$ and $\partial\alpha^{*UL}/\partial b > 0$, with strict inequalities holding for low enough marginal costs of gaming.

Existence of solutions with $\alpha \in (\alpha_L, \alpha_U)$

The first order condition (2) for solutions with $\alpha < \alpha_L$ implies that $N^{*L}(b) = (b - c_{N1})/c_{N2}$ for $b \geq b_1 = c_{P1}$. Now consider b' defined by $N^{*L}(b') = \alpha_L(P^o - E^o)$. We have

$$s_N^+ = b' - C_N = 0 < s_N^- = \overline{v}\pi'(P^o - E^o)^{-1} + b' - C_N \quad (21)$$

Hence there must exist a $b_2 < b'$ such that

$$b_2 - C_N(N^{*L}(b_2)) = 0 = b_2 + \overline{v}\pi'\left(P^*(b_2) - E^*(b_2)\right)^{-1} - C_N(N^{*LU}(b_2)) \quad (22)$$

and the practice is indifferent between the two solutions at b_2. For $b > b_2$ it strictly prefers solutions with $\alpha > \alpha_L$.

Solutions with $\alpha \in (\alpha_U, 1)$

Solutions above the upper threshold with $\alpha \in (\alpha_U, 1)$ satisfy

$$s_N = b - C_N = 0 \quad (23)$$

$$s_E^+ = -C_E^+ > 0 > -C_E^- \quad (24)$$

$$s_P = R_P - C_P = \overline{v}\pi^o F' - C_P = 0 \quad (25)$$

and $E^{*U} = E^o$, $P^{*U} > P^o$. Since N does not affect s_P increases in b or reductions in the marginal cost of N have no effect on P^* and E^* but do increase N^{*U} and α^{*U}. With sufficiently high b or low marginal cost parameters on N the solution $N^{*U} = P^{*U} - E^o$, $\alpha^{*U} = 1$ is obtained.

Solutions with $\alpha = \alpha_U$

The first order condition (5) on N for solutions with $\alpha \in (\alpha_L, \alpha_U)$ solves for

$$N^{*UL}(b) = \frac{R_N + b - c_{P1}}{c_{P2}} = \frac{\pi^o\left[(\alpha_U - \alpha_L)\left(P^{*UL}(b) - E^{*UL}(b)\right)\right]^{-1} + b - c_{P1}}{c_{P2}} \quad (26)$$

We know that $P^{*UL}(b)$ is decreasing and $E^{*UL}(b)$ increasing in b. Hence $N^{*UL}(b)$ increases at least linearly with b and so must $\alpha^{*UL}(b)$. Hence there exists $b_3 > b_2$ such that $\lim_{b^+ \to b_3} \alpha^{*UL}(b) = \alpha_U$ where limit is from below. The same conclusion holds if $E^{*UL}(b) = E^o$ or $P^{*UL}(b) = P^o$.

The first order condition on N for solutions with $\alpha \in (\alpha_U, 1)$ solves for

$$N^{*L}(b) = \frac{b - c_{P1}}{c_{P2}} \quad (27)$$

which is increasing linearly with b. So is $\alpha^{*U}(b)$ since E^{*U} and P^{*U} do not vary with b. Thus there exists a b_4 such that $\lim_{b^- \to b_4} \alpha^{*U}(b) = \alpha_U$ where the limit is from above. Since $P^{*UL}(b) \le P^o \le P^{*U}(b)$ and $E^{*UL}(b) \ge E^o = E^{*U}$ comparison of (26) and (27) shows that $b_4 > b_3$.

For solutions $b \in (b_3, b_4)$ where $\alpha = \alpha_U$, we substitute for $N = \alpha_U(P - E)$ in $s(N,E,P)$ and obtain the first order conditions

$$-C_E - \alpha_U(b - C_N) = 0$$

$$R_P - C_P + \alpha_U(b - C_N) = 0$$

when $E^{*\bar{U}}(b) > E^o$ and $P^{*\bar{U}}(b) < P^o$. Differentiation respect to b shows that $\partial E^{*\bar{U}}(b)/\partial b < 0$, $\partial P^{*\bar{U}}(b)/\partial b > 0$ and $\partial N^{*\bar{U}}(b)/\partial b > 0$. If solutions with $\alpha = \alpha_U$ have $E^{*\bar{U}}(b) > E^o$ and $P^{*\bar{U}}(b) = P^o$, then $\partial E^{*\bar{U}}(b)/\partial b < 0$ and $\partial N^{*\bar{U}}(b)/\partial b > 0$. Conversely if $E^{*\bar{U}}(b) = E^o$ and $P^{*\bar{U}}(b) < P^o$, then $\partial P^{*\bar{U}}(b)/\partial b > 0$ and $\partial N^{*\bar{U}}(b)/\partial b > 0$. Finally if $E^{*\bar{U}}(b) = E^o$ and $P^{*\bar{U}}(b) = P^o$, then $\partial N^{*\bar{U}}(b)/\partial b = 0$.

We summarise the comparative static properties of the model within solution types in the table.

Comparative statics

		Effect on			
Increase in	Solution type	*N*	*E*	*P*	α
Altruism *b* (or reduction in marginal cost of *N*)	$\alpha \in (\alpha_L, \alpha_U)$	+	+	–	+
	$\alpha > \alpha_U$	+	0	0	+
List size *M*; maximum points π^o	$\alpha \in (\alpha_L, \alpha_U)$	+	+	–	+
	$\alpha > \alpha_U$	0	0	+	–
True exceptions E^o	$\alpha \in (\alpha_L, \alpha_U)$	+	+	–	+
	$\alpha > \alpha_U$	0	0	0	0
True prevalence P^o	$\alpha \in (\alpha_L, \alpha_U)$	–	–	+	–
	$\alpha > \alpha_U$	0	0	+	–

Appendix D. Reported achievement and delivered quality

Determinants of reported achievement and delivered quality reported in Scottish practices 2005/6

	Reported achievement		Delivered quality	
	Coef	t	Coef	t
Prop ≤15	5.817	2.04	4.946	1.64
Prop >75	-9.406	-2.49	-0.935	-0.24
SMR	0.000	0.01	-0.001	-0.14
Income depriv	-0.090	-7.49	-0.117	-9.07
Ethnicity	-5.646	-1.07	-17.184	-2.98
Rural	0.104	0.41	0.850	3.20
Sparsity	0.010	0.72	0.032	2.10
List 1000	-0.137	-5.43	-0.112	-4.56
GP/list 1000	0.601	2.71	0.270	1.17
Prop female GP	0.396	1.21	0.527	1.57
Av GP age	-0.025	-1.54	-0.043	-2.65
Prop GP not UK qual	-1.252	-2.45	-1.004	-2.03
PMS contract	-0.445	-1.61	-0.881	-3.34
Dispensing	-0.404	-1.23	-0.165	-0.49
Training	-0.425	-2.70	-0.225	-1.35
Ex FH	0.711	5.17	0.161	1.14
New practice	0.206	0.70	-0.025	-0.09
Herfindahl	-1.615	-4.34	-2.045	-5.06
SPICE	0.872	6.33	0.507	3.56
Patient Inflow	-3.295	-1.62	-3.384	-1.88
Software: Egton	-0.354	-1.29	0.087	0.31
Software: In Practice	0.677	2.61	1.191	4.00
Software: Protechnic	1.102	2.22	1.068	1.56
Software: iSoft	0.121	0.32	0.542	1.30
Constant	92.874	66.46	88.908	62.36
$\overline{R}^2$		0.6227		0.7548
N		58453		56980

Achievement: 100*N/D. Delivered quality: 100*N/(D+E), where N is number of patients for whom indicator was achieved, D is the number declared suitable by the practice, E is the number exception reported. OLS regressions also include 14 Health Board dummies and 64 indicator dummies. Robust standard errors, adjusted for clustering in practices. *** p<0.001, **p<0.01, *p<0.05.

The standardised illness ratio (SIR), income deprivation and employment deprivation are highly correlated. We used income deprivation, and dropped SIR and other deprivation measures in the reported achievement and delivered quality models because of the considerable interest in income related inequity in health care delivery.