THE UNIVERSITY of York

CHE
Centre For Health Economics

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

CHE Research Paper 28

Doctor Behaviour Under a Pay for Performance Contract: 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 [a]
Health Economics Research Unit, University of Aberdeen[b]

May 2007

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, University of Bristol. We are grateful for comments made at presentations at CHE, CMPO, the Norwegian School of Economics and Business Administration, and the Royal Economics Society Conference, Warwick; and for discussions with Tim Doran, Bruce Guthrie, David Reeves and Martin Roland about the intricacies of the QOF.

* Correspondence to Hugh Gravelle: hg8@york.ac.uk

Disclaimer

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/inst/che/pubs. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321458 for further details.

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 declares 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 12.4% 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.

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.

JEL Nos: H51, I18

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

1. Introduction

In April 2004 the UK National Health Service (NHS) introduced what is possibly the most elaborate and expensive explicit pay for performance scheme in any health care system, or indeed in any organisation anywhere (Roland, 2004). All 10,000 general practices in the UK were required to report their achievements on 146 quality indicators in the Quality and Outcomes Framework (QOF). 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, and the amount paid out to practices under the QOF was around £1,000M in 2005/6.[1]

The QOF was not trialled before it was introduced simultaneously in all four parts of the UK. There are no national data on the performance of practices on the QOF indicators before the QOF was introduced. Consequently there is only fragmentary evidence on the effect of the QOF on incentivised and unincentivised activities. The tentative conclusion from studies based on limited sets of indicators in small samples of practices (Gulliford et al, 2006; Campbell et al, 2007), is that the quality of care in general practice had been increasing since the late 1990s, but that the QOF may have led to a further increase above this trend.

In this paper we construct a model of behaviour under the QOF by semi-altruistic GPs who care about their income and about patient welfare. We test the model using data from Scottish practices in 2004/5 and 2005/6.

Around half of potential QOF revenue is attached to indicators of clinical quality measured by 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. Practices are paid a price per point achieved and the number of points achieved increases linearly with the proportion treated between a lower threshold (0.25) and an upper threshold (0.70) (see Figure 1).

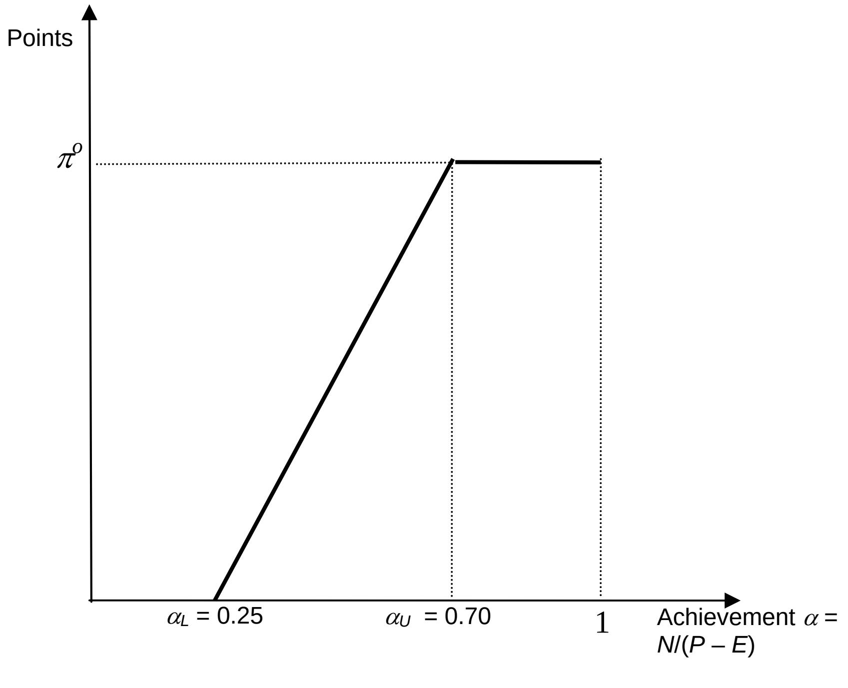

Figure 1. Achievement and points for a continuous quality indicator (CHD6)

Since increasing the number of patients treated is costly for GPs, only GPs willing to trade-off income for improved patient health will have performance above the upper threshold for a clinical indicator. Practice performance under the QOF is consistent with a substantial degree of altruism. In 2005/6 90.8% of practices were above the upper threshold for indicators and could, on average, have reduced the number of patients treated by 12.4% without affecting their QOF revenue.

[1] 2005/6 is the financial year 1 April 2005 to 31 March 2006. Similarly for the financial year 2004/5.

Expressing a clinical indicator as a ratio is intended to provide an incentive to GPs to increase the numerator, i.e. to increase the number of treated patients. But it is possible for GPs to change the denominator, i.e. the number they declare as eligible for treatment. The number declared eligible is the number of patients they report as having a condition minus the number they exception report for an indicator. It is possible to exception report patients as unsuitable for an indicator on a variety of grounds (Appendix A). For example, patients may be exception reported because they are terminally ill or because they have refused 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 rates are determined by the characteristics of the population served by the practice population. 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 that had previously held budgets under the fundholding regime. 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 for gaming uses the structure of the incentive scheme. The marginal rewards for overstating exceptions are positive when the indicator is below its upper threshold and zero when above the upper threshold. Similarly, practice revenue is increased by understating prevalence when it is below the upper threshold. Moreover, because of another feature of the scheme explained below, practice revenue is increased by overstating prevalence when the indicator is above the upper threshold. The price per point was 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 proportion treated was zero in both years. We argue that practices which were below the upper threshold in the first year of the scheme would therefore be more likely to increase their reported exception rate and reduce their reported prevalence in the second year. Practices already above the upper threshold would only have to maintain their first year behaviour to maximise their 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.

Section 2 sets out the QOF in more detail and provides a model of practice behaviour to guide the empirical work. Section 3 describes the data and the estimation methods. Section 4 has results and section 5 concludes with a discussion of the implications of the results.

2. Incentives in the QOF

2.1 GP 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 no charges for use of primary care[2] or for the rest of the NHS.

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

[2] Apart from a charge for prescription drugs. Because of the wide range of exemptions on grounds of age, income, and health, 92% of drug prescriptions dispensed were free in Scotland in 2005/6.
http://www.isdscotland.org/isd/info3.jsp;jsessionid=A436EB8C9E9ED916E8D06A4DDB99020C?pContentID=2237&p_applic=CCC&p_service=Content.show&

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. 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.[3]

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 Table 1. There are additional points for holistic care (determined by the 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.

Table 1. 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

There are clinical quality indicators in 11 disease domains (Table 2). We concentrate on the 65 continuous 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 (see 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 the disease domain.

[3] PMS practices had a deduction of 168 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)

Table 2. Disease domains: continuous 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	0.78	0.41	0.93	0.15	5.43	1.09
Cancer	1	6	0.69	0.46	0.92	0.18	0.71	0.23
CHD	11	95	0.79	0.40	0.95	0.11	4.54	1.16
COPD	7	40	0.66	0.47	0.89	0.19	1.84	0.93
Diabetes	17	93	0.75	0.43	0.91	0.13	3.40	0.72
Epilepsy	3	14	0.65	0.48	0.83	0.24	0.72	0.21
Hypertension	4	96	0.75	0.43	0.91	0.21	12.08	2.81
Hypothyroidism	1	6	0.91	0.28	0.98	0.13	2.99	0.94
LVD	2	16	0.78	0.42	0.93	0.16	0.63	0.28
Mental health	4	34	0.76	0.43	0.92	0.18	0.59	0.32
Stroke and TIA	9	27	0.75	0.44	0.94	0.13	1.90	0.57
Total	65	492						

Prevalence rate: List weighted mean of raw prevalence rates. Upper threshold achievement (*dom upper*): average proportion of indicators in domain for which upper threshold achieved weighted by maximum possible points for domain. Figures are weighted by listsize. CHD: coronary heart disease. COPD: chronic obstructive pulmonary disease. LVD: left ventricular dysfunction. TIA: transient ischaemic attacks.

* Mean and SD of achievement rates exclude the first indicator of all disease domain, which ask practices whether or not they have a register of patients recorded as having such disease.

Table 3. 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

The achievement by practice g on a continuous indicator i in disease domain k is $\alpha_{gki} = N_{gki} / D_{gki} = N_{gki}/(P_{gk} - E_{gki})$. (Table 4 summarises the notation.) N_{gki} is the number of patients for whom indicator i in domain k for whom the indicator is achieved eg the number with CHD whose blood pressure has been measured in the last 15 months (CHD5). $D_{gki} = P_{gk} - E_{gki}$ is the number of patients declared as suitable for indicator i in domain k. Prevalence P_{gk} is the number of patients in disease domain k reported by practice g. E_{gki} 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). 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.

Table 4. 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.

The number of points (and hence revenue) earned on an indicator varies linearly with achievement for $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$. See Figure 1. No points are earned if α is less than or equal to the lower threshold α_{kiL} (25% for all continuous indicators). At the upper threshold α_{kiU} the practice receives the maximum number of points π^o_{ki} 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 5).

Table 5. Summary statistics: rates of achievement, delivered quality, exception reporting for 41 indicators. Scotland 2005/6

Indicator	Maximum points available	Upper threshold (%)	% achieving upper threshold	Reported achievement (%)		Delivered quality		Exception reporting (%)	
				Mean	SD	Mean	SD	Mean	SD
ASTHMA6	20	70	89.7	80.9	11.0	74.1	11.1	8.8	8.9
BP2	10	90	98.9	98.2	1.8	97.1	2.2	1.2	1.1
BP4	20	90	83.6	93.4	4.5	91.4	4.6	2.2	2.5
BP5	56	70	89.5	79.1	7.2	74.6	7.5	5.8	4.0
CHD3	7	90	97.5	97.4	2.7	95.7	3.0	1.8	1.7
CHD5	7	90	97.7	97.5	2.5	95.5	3.0	2.2	2.1
CHD6	19	70	99.3	89.2	5.3	85.3	5.6	4.5	3.2
CHD7	7	90	88.4	94.3	4.8	90.4	5.2	4.3	3.4
CHD8	16	60	95.8	80.5	8.7	72.1	7.9	10.7	5.5
CHD9	7	90	94.0	95.3	3.3	92.2	3.6	3.4	2.6
CHD10	7	50	99.2	75.5	11.6	55.0	6.4	27.5	11.1
CHD12	7	85	93.9	93.0	4.9	80.4	5.9	13.7	6.0
COPD3	5	90	80.7	92.6	12.2	82.8	12.9	11	8.8
COPD4	6	90	95.6	96.9	3.7	94.0	4.8	3.2	3.3
COPD6	6	70	90.3	88.7	14.6	77.4	14.8	13.0	10.1
COPD8	6	85	94.7	93.6	5.4	80.8	7.3	14.3	7.0
DM2	3	90	94.1	95.9	3.4	92.5	4.1	3.7	2.8
DM3	3	90	99.3	98.5	1.8	97.0	2.5	1.5	1.6
DM5	3	90	98.8	97.8	2.1	94.9	3.0	3.1	2.3
DM6	16	50	85.5	59.7	10.5	51.5	9.8	13.5	8.5
DM7	11	85	94.4	92.2	3.8	86.4	4.8	6.4	3.8
DM8	5	90	76.8	92.7	7.2	86.2	7.7	7.3	5.4
DM9	3	90	69.0	90.6	7.6	84.8	8.6	6.5	5.0
DM10	3	90	67.0	89.8	8.2	84.1	9.0	6.7	5.1
DM11	3	90	99.5	98.7	1.5	96.9	2.1	1.9	1.7
DM12	17	55	98.5	78.9	8.5	73.0	9.0	7.7	4.5
DM14	3	90	97.2	96.9	3.3	94.0	4.7	3.0	2.8
DM16	3	90	96.9	96.7	2.6	93.8	3.5	3.0	2.4
DM17	6	60	98.1	81.3	7.5	72.7	7.5	10.4	4.9
DM18	3	85	91.3	92.0	5.5	77.1	6.6	16.3	6.8
LVD3	10	70	95.1	86.9	8.3	57.0	12.7	20.9	14.1
EPILEPSY4	6	70	66.8	72.7	13.5	78.8	9.3	9.5	8.7
MH2	23	90	91.0	96.3	4.6	90.5	10.3	6.1	9.1
MH5	5	70	80.4	87.7	21.6	69.8	31.7	10	4.6
STROKE3	3	90	95.1	96.3	3.4	93.5	4.4	3.1	3.1
STROKE5	2	90	96.8	96.9	3.0	93.6	4.4	3.5	3.8
STROKE6	5	70	98.1	87.8	6.0	81.9	7	6.9	5.1
STROKE7	2	90	80.4	92.4	6.4	86.1	7.5	7	5.8
STROKE8	5	60	92.5	76.7	9.9	65.4	9.5	14.8	7.9
STROKE10	2	85	85.6	90.5	6.4	75.3	7.6	16.8	7.6
THYROID2	6	90	96.9	96.9	3.0	96.0	3.1	0.9	1.3

Reported achievement: N/D; delivered quality: $N/(D+E)$; Exception reporting: $E/(D+E)$. Denominator weighted means over Scottish practices.

Formally, the number of points earned by practice *g* for indicator *i* in disease domain *k* is

$$\pi_{gki} = \pi_{ki}(\alpha_{gki}) = \pi^{o}_{ki} \min\{1, \max\{(\alpha_{gki} - \alpha_{kiL})/(\alpha_{kiU} - \alpha_{kiL}), 0\}\} \quad (1)$$

The revenue from indicator *i* for condition *k* is the price per point times the number of points

$$R_{gki} = v_{gk}\pi_{ki}(\alpha_{gki}) \quad (2)$$

The price per point

$$v_{gk} = vF_{gk}\frac{M_g}{\bar{M}} \quad (3)$$

is the product of the national average price per point *v*, the adjusted practice disease prevalence factor F_{gk} for disease *k* for practice *g*, and the list size of the practice relative to the national average list size $M_g/\bar{M}$, where $\bar{M} = \sum_{g'} M_{g'}/G$ is average list size and *G* is the number of practices. The national average price per point *v* was £70 in 2004/5 and £124.60 in 2005/6.

The adjusted disease prevalence factor 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:[4]

$$F_{gk} = F_{gk}(P_{gk}) = \left(P_{gk}/M_g\right)^{\frac{1}{2}}\left[G^{-1}\sum_{h}^{G}\left(P_{hk}/M_h\right)^{\frac{1}{2}}\right]^{-1}, \qquad F'_{gk}(P_{gk}) > 0 \quad (4)$$

Thus practice *g* revenue from indicator *i* in disease domain *k* is

$$\begin{aligned} R_{gki} &= v\left[M_g/\bar{M}\right]\pi_{ki}(\alpha_{gki})\left[P_{gk}/M_g\right]^{\frac{1}{2}}\left[G^{-1}\sum_h (P_{hk}/M_h)^{\frac{1}{2}}\right]^{-1} \\ &= \bar{v}_g\pi_{ki}(\alpha_{gki})F_{gk}(P_{gk}) \end{aligned} \quad (5)$$

Notice that, provided $\alpha_{gki} > \alpha_{kiL}$, practice revenue increases with the square root of the list size M_g. Guthrie, McLean and Sutton (2006) show that this can lead to large variations in QOF income for practices with the same reported prevalence and the same achievement. Moreover, provided $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$, the marginal revenue from increasing achievement α_{gki} also increases with the square root of the practice list. Hence practices with more patients have a greater incentive to increase *N*, *E* and *P*.

A practice of given size can alter its QOF revenue from a continuous indicator by changing

- the number of patients for whom the indicator is met (N_{gki})
- the number of patients who are exception reported for that indicator (E_{gki})
- the reported number of patients with the relevant condition (P_{gk})

When $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$ increasing the number of patients for whom the indicator is met (N_{gki}) or increasing exception reporting (E_{gki}) will increase revenue.

The effect of reporting a larger number of prevalent patients (P_{gk}) is more complicated. An increase in P_{gk} reduces achievement $\alpha_{gki} = N_{gki}/(P_{gk} - E_{gki})$ and increases the adjusted disease prevalence factor $F_{gk}(P_{gk})$. If $\alpha_{gki} < \alpha_{kiL}$ then $\pi'_{ki}(\alpha_{gki}) = 0$ and $\pi_{ki}(\alpha_{gki}) = 0$, so that $\partial R_{gki}/\partial P_{gk} = 0$. When $\alpha_{gki} > \alpha_{kiU}$ we have $\pi'_{ki}(\alpha_{gki}) = 0$ and $\pi_{ki}(\alpha_{gki}) > 0$, so that $\partial R_{gki}/\partial P_{gk} > 0$ because of the increase in the adjusted disease prevalence factor. In the intermediate range $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$ increases in P_{gk} reduce achievement α_{gki} but increase the adjusted disease prevalence factor.

---

[4] We ignore the additional complication that the practice prevalence rate is truncated at the 5th centile both in calculating the practice rate and the national average rate.

We can show (see Appendix) that the overall effect is to reduce revenue from the indicator. Thus

$$\frac{\partial R_{gki}}{\partial P_{gk}} = \bar{v}_g \pi'_{gk}(\alpha_{gki}) \frac{\partial \alpha_{gki}}{\partial P_{gk}} F_{gk}(P_{gk}) + \bar{v}_g \pi(\alpha) F'_{gk}(P_{gk})$$

$$= 0, \qquad \alpha_{gki} < \alpha_{kiL}$$

$$< 0, \qquad \alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$$

$$> 0, \qquad \alpha_{gki} > \alpha_{kiU} \qquad (6)$$

with marginal revenue jumping down from zero at α_{kiL} and jumping up from negative to positive at α_{kiU}. 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.

Table 6 summarises the properties of the indicator revenue function.

Table 6. Indicator revenue function: partials and cross partial derivatives

		Effect on			
		R_{gki}	$\partial R_{gki} / \partial N_{gki}$	$\partial R_{gki} / \partial E_{gki}$	$\partial R_{gki} / \partial P_{gki}$
Change in	Achievement level	(1)	(2)	(3)	(4)
N_{gki}	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	+	0	+	–
	$\alpha_{gki} > \alpha_{kiU}$	0	0	0	0
E_{gki}	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	+	+	+	–
	$\alpha_{gki} > \alpha_{kiU}$	0	0	0	0
P_{gk}	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	–	–	–	–
	$\alpha_{gki} > \alpha_{kiU}$	+	0	0	–
M_g	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	+	+	+	–
	$\alpha_{gki} > \alpha_{kiU}$	+	0	0	+
α_{kiu}	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	–	+	+	–
	$\alpha_{gki} > \alpha_{kiU}$	0	0	0	0
π^o_{ki}	$\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$	+	+	+	–
	$\alpha_{gki} > \alpha_{kiU}$	+	0	0	+

R_{gki} revenue from indicator i in disease domain k. N_{gki} number patients indicator achieved for. E_{gki} number exception reported. P_{gk} declared prevalence in disease domain. $\alpha_{gki} = N_{gki}/(P_{gk} – E_{gki})$ achievement ratio. α_{kiL} lower threshold. α_{kiU} upper threshold. π^o_{ki} maximum points for indicator. M_g practice population. The effect of practice population M_g is via its effect on the value of a point v_{gk} and the effects of α_{kiu} and π^o_{ki} are via the number of points achieved $\pi_{ki}(\alpha_{gki})$.

2.3 A model of the QOF

2.3.1 Optimal treating and cheating

To derive hypotheses to structure the empirical analysis we assume that practice utility is linear in patient health and practice income (QOF revenue minus costs) and that there is only one continuous clinical quality indicator:

$$\begin{aligned} s(N,E,P) &= bN + R(N,E,P) - C(N,E,P) \\ &= bN + \bar{v}\pi(\alpha)F(P) \; - c_{N1}N - c_{E1}\left(\left|E-E^o\right|/P^o\right) - c_{P1}\left(\left|P-P^o\right|/M\right) \\ &\quad - \tfrac{1}{2}\left[c_{N2}N^2 + c_{E2}\left((E-E^o)/P^o\right)^2 + c_{P2}\left((P-P^o)/M\right)^2\right] \end{aligned} \quad (7)$$

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 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 [5]

$$\frac{\partial s}{\partial N} = b + \bar{v}\pi'(\alpha)\frac{1}{P-E}F(P) - c_{N1} - c_{N2}N \quad (8)$$

$$\frac{\partial s}{\partial E} = \bar{v}\pi'(\alpha)\frac{N}{(P-E)^2}F(P) - c_{E1}\frac{d\left|E-E^o\right|}{dE}\frac{1}{P^o} - c_{E2}\frac{(E-E^o)}{P^{o2}} \quad (9)$$

$$\begin{aligned} \frac{\partial s}{\partial P} &= -\bar{v}\pi'(\alpha)\frac{N}{(P-E)^2}F(P) + \bar{v}\pi(\alpha)F'(P) \\ &\quad - c_{P1}\frac{d\left|P-P^o\right|}{dP}\frac{1}{M} - c_{P2}\frac{(P-P^o)}{M^2} \end{aligned} \quad (10)$$

---

[5] 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).

The seven possible solutions to the problem of maximising (7) subject to non-negativity constraints on *N*, *E*, and *P* and to $\alpha = N/(P-E) \in [0,1]$ are illustrated in Figure 2.[6]

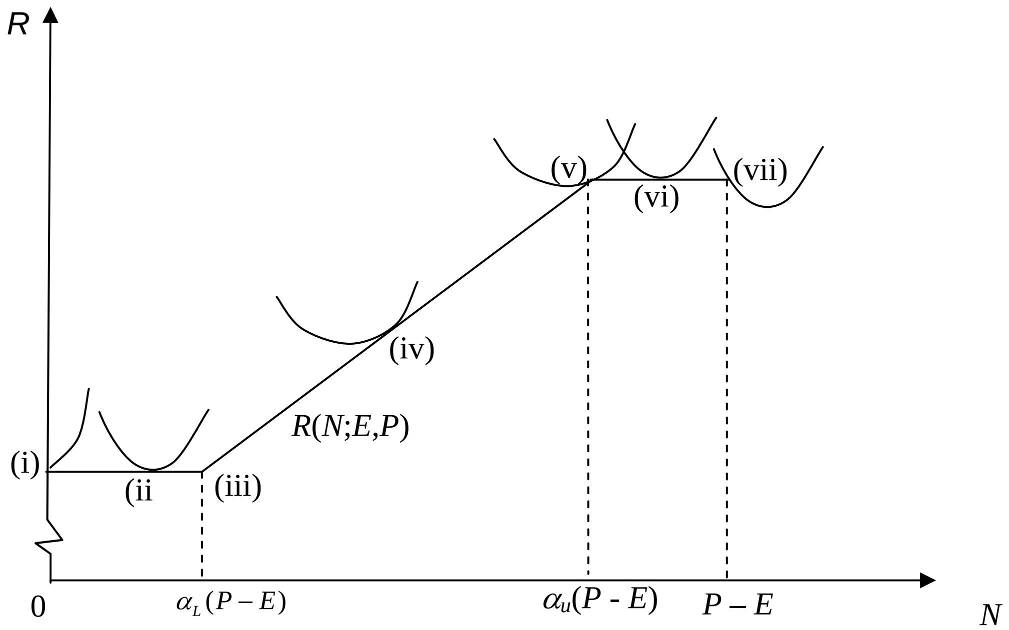

Figure 2. Solution types

(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. 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 $N = P - E^o$.

[6] 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 (5). The marginal utility of treatment is $S_N = b - c_N N$.

The solution types are summarised in Figure 3.

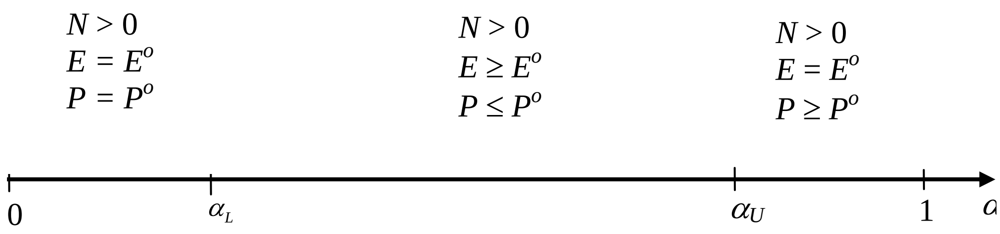

Figure 3. Solution types and achievement level

2.3.2 Testing for gaming

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 for details). 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*. We can establish (see Appendix)

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 can use the proposition to derive tests for gaming of prevalence and exceptions.

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 $\alpha > \alpha_U$. Hence, other things than *b* held constant, practices with $\alpha \in (\alpha_L, \alpha_U)$ will have declared prevalent prevalence rates *P*/*M* which are not larger than those with $\alpha > \alpha_U$. For some sets of cost parameters, with low marginal costs of gaming, the prevalence rate for practices with $\alpha \in (\alpha_L, \alpha_U)$ will be less than those with $\alpha > \alpha_U$. Thus a comparison of the declared prevalence rates of practices above and below the upper threshold with similar observable characteristics affecting cost parameters, true prevalence, and true exceptions would enables us to test for gaming of prevalence.

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 have 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.

Since the new 2004/5 contract embodying the QOF was a radical departure from the previous contract it is plausible that practices in 2004/5 would have been uncertain about the performance they would achieve. It is plausible that they would use 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 income in 2005/6. Those above the upper threshold would know that they could only increase income by increasing prevalence. Hence we test for gaming of prevalence reporting by examining whether practices which were below the upper threshold in 2004/5 have a lower reported prevalence in 2005/6 than those which were above the upper threshold in 2004/5.

Exception reporting above and below the upper threshold

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 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 cross sectional comparison of exception reporting of practices above and below the threshold in a year since unobserved factors increasing exceptions will increase the likelihood that a practice is above the upper threshold. Hence the test for gaming is biased towards rejection of gaming. Hence we again use the fact that the practice was above or below the upper threshold in 2004/5 to test for exception gaming in 2005/6. 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.

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 whether they hold a GMS or PMS contract. Gamed prevalence and exceptions will vary with practice characteristics that reflect GP cost or preference parameters. 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 population morbidity, deprivation, rurality, and ethnicity to reduce this omitted variable problem.

3. Data and methods

3.1 Data

We used QOF data on Scottish practices because it was possible to link them to more practice level characteristics than for practices in other parts of the United Kingdom. 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 Information Services Division (ISD) of NHS National Services Scotland.[7] Only practice-level counts are available and, for confidentiality, these are suppressed if they take a value less than 5. We discuss the measurement of prevalence and exception reporting in section 3.2.

Practice characteristics such as the proportion of female GPs, the average age of GPs, the type of contract (GMS or PMS), list size, number of GPs etc were provided from the GP Contractor Database held at ISD and are for 1 April 2005. ISD also provided information on whether the practice had previously held a budget under the fundholding scheme (Dusheiko et al, 2006) which was in operation between 1991/2 and 1998/9.

We calculated a 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 calculating 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.

[7] www.isdscotland.org/QOF

Practice population morbidity was measured by the Standardised Illness Ratio (SIR.) This is an indirectly age standardised measure of the proportion of people reporting a limiting long-term illness at the 2001 Census. The Census also provided the proportion of residents in black and ethnic minority groups. Census data are available for each of 42,604 Census Output Areas containing on average 117 individuals in private households.

Table 7. Summary statistics: explanatory variables

Variable	Definition	Mean	SD	Min	Max
Patient characteristics					
SIR	Standardised Illness Ratio	98.446	22.670	50.533	186.742
SMR	Standardised Mortality Ratio	107.457	25.135	44.199	329.250
Prop15less	Proportion of patients aged 15 or less	0.164	0.029	0.016	0.270
Prop75plus	Proportion of patients over 75 years	0.070	0.021	0.001	0.159
Inc	Income deprivation score	14.962	7.583	2.801	43.409
Educ	Education deprivation score	-0.007	0.591	-1.583	1.671
Ethnicity	Minority ethnic group proportion	0.021	0.028	0.000	0.408
Rural	Population mode in settlements <3,000 people	0.107	0.310	0.000	1.000
Sparsity	Inverse of population density (hectares per person)	1.516	5.470	0.014	136.375
Practice characteristics					
Pop1000	List size in 000s	5.248	3.259	0.118	23.097
Gp1000	GPs per 1,000 patients	1.039	0.913	0.263	13.793
Pfemale	Female GPs	0.423	0.255	0.000	1.000
Avgpage	Average age of GPs (years)	44.913	6.123	20.000	67.000
Pcqnonuk	GPs not qualified in UK	0.099	0.225	0.000	1.000
Pmscont	PMS contract	0.105	0.306	0.000	1.000
Dispensing	Dispensing practice	0.136	0.343	0.000	1.000
Exfh	Ex fundholding practice	0.491	0.500	0.000	1.000
Herf	Herfindahl index for areas served	0.366	0.251	0.078	0.998

Based on 916 practices. One observation per practice. Patient characteristics are weighted by list size. Practice characteristics are unweighted.

We also measured population health with the Standardised Mortality Ratio (SMR) calculated by ISD for each data zone using General Register Office death records for 1999 to 2003 and the 2001 Census population counts.

Socio-economic factors were taken from the Scottish Index of Multiple Deprivation 2004. The Index consists of an overall deprivation score calculated from deprivation scores for six deprivation domains: income; employment; housing; education, skills and training; health and disability; and access to services. The overall deprivation score and the deprivation domain scores are available at data zone level.

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.

Summary statistics for the population and practice characteristics are given in Table 7.

3.2 Measurement of QOF variables

We estimate models for achievement and delivered quality, and 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

$$D_{gki} + E_{gki} = P_{gk} + A_{gk} = P_{gk}^{T} \quad (11)$$

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 $P_{gk}^{T} = D_{gki} + E_{gki}$ for 2005/6. We measure the exception reporting rate for indicator *i* in disease domain *k* as $E_{gki}/(D_{gki} + E_{gki})$.

We expect that the age and sex mix of the practice population will affect reported prevalence P_{gk} since the age and sex mix is the main determinant of true prevalence. We could allow for demographics by including the proportions of the practice population as explanatory variables. This would have a cost in terms of degrees of freedom and 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.[8] The indirectly standardised reported prevalence ratio was obtained by dividing the observed prevalence reported in QOF by expected prevalence.[9] We report prevalence regressions for 2005/6 in eight of the eleven disease domains because of the lack of information on (similarly defined) 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 ($p<0.001$).

Set of indicators used

Some of the clinical indicators for a disease domain are applicable only to subsets of the prevalent population with that condition. For example, for asthma, the prevalence report is the number of patients with asthma but indicator ASTHMA7 is the proportion of patients with asthma over the age of 16 who have had influenza immunisation in the preceding 1 September to 31 March. We limit our analysis to 41 indicators where the target population is the whole population with the condition, not a subset of it.

Reported and delivered quality

We investigate the determinants of quality of care provided in practices. Practices are rewarded for reported quality: the number (*N*) for whom an indicator has been achieved divided by the number of patients with the condition who are reported as suitable for the treatment (*D*). Since practices may exception report patients for an indicator we prefer to measure quality by reference to the number of

---

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

[9] $P_{gk}/\sum_a p^o_{ka}\omega_{ag} = \sum_a p_{kag}\omega_{ag} / \sum_a p^o_{ka}\omega_{ga}$ where p^o_{ka} 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*.

patients with the condition who could have benefited from a particular treatment. We therefore define the quality delivered to patients as the ratio of number treated to the number who could have been treated: *N*/(*D*+*E*).

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_{gki} (*ind upper*) taking the value 1 if practice *g* had reported achievement (N_{gki}/D_{gki}) above the upper threshold α_{kiU} for indicator *i* in disease domain *k*.

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*: *dom upper*$_{gk}$ = $\sum_i \pi^o_{ki} U_{gki} / \sum_i \pi^o_{ki}$ where π^o_{ki} is the maximum number of points available for indicator *i* in disease domain *k*.

3.3 Estimation methods

The models of practice performance and reporting of prevalence are estimated by ordinary least squares. We estimated the exception reporting model by both ordinary least squares and by negative binomial regression using *D*+*E* as the exposure term. In all models allowance is made for clustering within practices for prevalence models or within practices and disease domains for models of performance and exception reporting. The models also include 14 Health Board dummy variables to allow for unobserved area effects and local policies affecting the delivery of primary care.

4. 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. Table 2 shows the high and increasing proportion of Scottish practices achieving the upper thresholds (and hence maximum points for indicators) for the different disease domains.

Table 5 shows, for 41 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*). The percentage achieving the threshold varied from 67.0 (DM10) to 99.3 (DM3). Delivered quality is lower than reported achievement, by virtue of its definition, but is also more variable across practices. Exception reporting ranges from 0.9% (THYROID2) to 27.5% (CHD10).

Many practices treated more patients than necessary to achieve the maximum points on indicators. Some 90.8% had reported achievement above the upper threshold. They could have reduced the number treated by 12.4% without reducing their QOF revenue. We interpret this substantial over achievement as evidence of GP altruism: practices are motivated by patient welfare as well as GP income.[10]

4.1 Reported achievement and delivered quality

Table 8 reports the regression models examining the determinants of reported achievement (N/D) and delivered quality (N/(D+E)). Delivered quality is lower in practices with more income deprived patients,

[10] The averages are weighted by the maximum points achievable for each indicator. Thus the percentage above the upper threshold is $100 \times \sum_g \left(\sum_i U_{gi} \pi^o_i \Big/ \sum_i \pi^o_i \right) G^{-1}$ where U_{gi} = 1 if practice *g* is above the upper threshold for indicator *i*, π^o_i is the maximum points achievable for indicator *i* and *G* the number of practices. The average reduction in treatment possible without reducing QOF revenue is $100 \times \sum_g \left(\sum_i U_{gi} (\alpha_{gi} - \alpha_{iU}) \alpha^{-1}_{gi} \pi^o_i \Big/ \sum_i \pi^o_i \right) G^{-1}$ where α_{gi} is achievement by practice *g* for indicator *i* (number treated divided by number declared eligible for treatment) and α_{iU} is the upper threshold for indicator *i*.

more ethnic minority patients, and higher in rural and more sparsely populated areas. This suggests that the quality gains arising from the QOF are delivered inequitably with respect to income and ethnicity.

Table 8. Determinants of reported achievement and delivered quality in Scottish practices 2005/6: pooled regressions

	Reported achievement		Delivered quality	
	Coef	t	coef	t
SMR	-0.004	-1.01	-0.005	-1.13
Prop15less	4.267	1.47	2.883	0.91
Prop75plus	-12.238	-3.20	-6.188	-1.53
Inc	-0.041	-1.74	-0.120	-4.55
Educ	-0.524	-1.88	0.091	0.30
Ethnicity	-10.475	-2.01	-21.532	-3.71
Rural	-0.122	-0.46	0.743	2.62
Sparsity	0.014	1.12	0.028	2.18
Pop1000	-0.147	-5.66	-0.111	-4.29
Gp1000	0.205	0.98	-0.041	-0.19
Pfemale	0.756	2.23	0.905	2.56
Avgpage	-0.015	-1.03	-0.042	-2.86
Pcqnonuk	-0.123	-0.31	-0.085	-0.20
Pmscont	-0.366	-1.31	-0.939	-3.34
Dispensing	-0.754	-2.31	-0.447	-1.28
Exfh	0.594	4.39	0.114	0.80
Herf	-1.446	-3.74	-2.022	-4.75
Constant	85.095	63.56	80.005	55.96
$\bar{R}^2$		0.6667		0.7729
N		35843		34705

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 40 indicator dummies. Robust standard errors, adjusted for clustering in practices. *** p<0.001, ** p<0.01, * p<0.05.

Delivered quality was higher in practices with a higher proportion of female GPs and younger GPs. Larger practices have worse delivered quality as do those exposed to less competition as measured by the practice Herfindahl index. Practices with a PMS contract also have lower QOF quality but this may be due to the fact that they were improving other aspects of care by agreement with their local Health Boards.

The differences between the effect of the explanatory variables on reported achievement and delivered quality are due to the way these variables affect exception reporting. Thus for example, being an ex fundholder improves reported achievement and has no effect on delivered quality, suggesting that ex fundholders exception reported more patients. Our exception reporting models provide more direct evidence of this.

4.2 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 9 reports two models in which prevalence reporting is pooled across 8 disease domains. The results are very similar for most of the coefficients and generally plausible. Reported prevalence is higher in practices with populations with higher standardised illness ratios and with more education deprivation.[11]

[11] The standardised illness ratio (SIR), income deprivation and employment deprivation are extremely highly correlated. We used income deprivation, and dropped SIR and employment deprivation in the reported achievement and delivered quality models because of the considerable interest in income related inequity in health care delivery (Wagstaff and van Doorslaer,

Reported prevalence is higher in practices with a higher proportion of female GPs and younger GPs. It is also higher in practices with more GPs per patient, suggesting that such practices have a lower cost of case finding.

Table 9 Determinants of prevalence reporting in Scottish practices 2005/6: pooled models

Dependent variable	Standardised reported prevalence rate 2005/6[1]			
	Coef.	t	Coef.	t
dom upper	-4.418	-1.62		
dom upper_lag			4.060	3.02
Sir	0.392	5.85	0.398	5.96
SMR	0.011	0.43	0.014	0.56
Educ	8.517	4.07	8.510	4.09
Ethnicity	-90.071	-3.60	-100.854	-3.11
Rural	-1.743	-0.96	-1.934	-1.07
Sparsity	-0.318	-4.93	-0.329	-5.14
pop1000	-0.426	-1.71	-0.414	-1.70
gp1000	5.677	4.41	5.581	4.32
Pfemale	3.717	1.81	3.512	1.73
Avgpage	-0.291	-3.54	-0.284	-3.50
Pcqnonuk	1.605	0.58	1.347	0.48
Pmscont	-0.005	0.00	1.010	0.69
Dispensing	-0.187	-0.11	0.902	0.58
Exfh	0.787	0.98	0.486	0.60
Herf	6.604	2.71	7.131	2.97
Constant	73.135	7.76	64.842	7.19
$\bar{R}^2$	0.2325		0.2354	
N	7318		7214	

[1] Age and sex indirectly standardised disease reported prevalence.
t stats are robust and allow for clustering within practices. All models are OLS and include 14 NHS Board dummies.
dom upper: proportion of indicators in domain for which practice achieved upper threshold in 2005/6. weighted by max points for indicators
dom upper_lag: proportion of indicators in domain for which practice achieved upper threshold in 2004/5 weighted by max points for indicators.

Model 1 includes a variable measuring the average extent to which the practice was above the upper threshold for the indicators in the disease domain (*dom upper*). It has a negative and insignificant coefficient. But whether the practice is above or below the upper threshold is endogenous, so that the coefficients on our measure of the practice being above or below the upper threshold (*dom upper*) will be biased. If an unobserved factor increases prevalence the practice is more likely, ceteris paribus, to be below the upper threshold. Hence the coefficients on *dom upper* will be biased downwards.

Model 2 uses *lag dom upper*, measuring the extent to which the practice was above the upper threshold in the previous year. The coefficient on the *lag dom upper* is positive and significant, suggesting that practices game prevalence reporting.

The pooled model forces the effects of the covariates to be the same for all disease domains. Given the differences in aetiology this is implausible. We therefore report results from separate models for standardised prevalence reporting in 2005/6 for eight disease domains in Table 10. There are noticeable differences in the effects of explanatory variables across the domains. The most striking is the effect of ethnicity, which is negative in seven of the domains but positive in the case of diabetes, a condition where ethnic minority patients are known to be at higher risk.

2000). However, for the models investigating exception and prevalence reporting it is more important to attempt to control for factors directly affecting health and so we include SIR, omitting income and employment deprivation.

Table 10 Determinants of prevalence reporting in Scottish practices 2005/6

Disease:	CHD	Diabetes	Asthma	COPD	Epilepsy	BP	Thyroid	Stroke
dom upper lag	3.007	10.384	7.491	0.046	1.961	3.643	4.793	9.182
	1.164	4.217	2.815	0.011	0.710	2.055	1.162	2.468
Sir	0.451	0.202	0.105	1.363	0.881	-0.002	-0.005	0.226
	5.191	2.769	1.059	6.120	5.905	-0.021	-0.055	2.046
SMR	0.041	0.075	0.007	-0.029	-0.031	0.006	0.018	0.026
	1.34	3.029	0.239	-0.414	-0.773	0.187	0.331	0.609
Educ	9.209	17.012	5.362	18.204	-4.372	9.656	1.464	10.336
	3.39	7.06	1.692	2.56	-0.901	3.74	0.479	2.897
Ethnicity	-108.923	115.344	-83.41	-189.902	-289.743	-68.023	-74.551	-103.521
	-2.633	1.529	-2.547	-2.512	-5.413	-1.849	-1.564	-3.011
Rural	-3.108	-1.73	-3.492	1.531	-2.209	-2.844	-2.138	-1.722
	-1.234	-0.924	-1.347	0.374	-0.665	-1.18	-0.651	-0.48
Sparsity	-0.179	-0.154	-0.484	-0.381	-0.673	-0.199	-0.33	-0.231
	-2.056	-2.019	-3.68	-2.189	-4.204	-2.316	-2.626	-1.869
pop1000	-0.054	-0.511	-0.083	-0.784	-0.202	-0.73	-0.832	-0.08
	-0.145	-2.26	-0.304	-1.656	-0.681	-2.419	-2.59	-0.198
gp1000	-0.368	0.986	7.706	10.781	15.005	2.658	-1.145	9.628
	-0.211	0.668	3.254	2.674	4.517	1.465	-0.404	3.878
Pfemale	0.294	0.149	2.036	10.714	-0.193	2.858	7.646	4.255
	0.103	0.06	0.578	1.475	-0.046	0.997	1.881	1.083
avgpage	-0.239	-0.218	-0.376	-0.415	-0.276	-0.089	-0.107	-0.514
	-2.362	-2.278	-2.712	-1.649	-1.492	-0.739	-0.609	-3.125
pcqnonuk	2.395	2.813	1.624	8.811	-7.021	5.536	-1.676	-1.775
	0.64	0.809	0.417	1.089	-1.207	1.475	-0.332	-0.353
pmscont	1.634	0.149	1.785	4.446	1.073	-0.757	-4.295	4.096
	0.956	0.086	0.619	0.925	0.433	-0.328	-2.037	1.329
dispensing	3.058	3.996	4.809	-2.285	-1.323	1.16	5.735	-7.157
	1.387	2.349	1.692	-0.638	-0.28	0.468	1.186	-2.382
Exfh	0.345	0.099	0.149	5.488	-1.204	-1.911	-2.091	3.057
	0.311	0.096	0.107	2.116	-0.694	-1.559	-1.346	1.895
Herf	2.558	13.258	9.473	7.516	-9.872	10.822	12.19	11.138
	0.699	4.09	2.455	1.118	-2.121	3.052	2.983	2.331
constant	58.535	61.601	86.44	-9.515	26.992	105.239	89.728	85.911
	5.094	5.311	6.272	-0.334	1.397	8.776	5.934	5.377
$\bar{R}^2$	0.6020	0.5694	0.1511	0.6218	0.4210	0.2032	0.5190	0.3176
N	903	902	903	902	895	903	903	903

Dependent variables: age and sex indirectly standardised disease reported prevalence 2005/6. All models included 14 NHS Board dummies. OLS coefficient reported above robust t statistic adjusted for clustering within practices.

The effect of *upper dom lag* is positive in all domains and is significant in four of them, suggesting that practices that had been above the upper thresholds in the domain in 2004/5 gamed their reporting of prevalence upward compared with those who were below the upper thresholds.

4.3 Exception reporting

Table 11 reports regressions of exception reporting pooled over 41 indicators. In all models exception reporting is 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. Practices that were fundholding have higher exception reporting.

Table 11. 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
Ind upper	0.786	4.92			0.145	6.67		
Lag_ind upper			-0.865	-7.60			-0.151	-9.42
Sir	0.032	3.69	0.028	3.25	0.003	2.64	0.003	2.23
SMR	0.004	1.27	0.005	1.34	0.001	2.20	0.001	2.23
Prop15less	1.526	0.59	1.751	0.68	-0.265	-0.71	-0.171	-0.46
Prop75plus	-4.829	-1.56	-6.054	-1.96	-0.824	-1.65	-0.912	-1.83
Educ	-0.674	-2.42	-0.636	-2.31	-0.105	-2.41	-0.102	-2.36
Ethnicity	15.925	4.99	15.298	4.88	1.892	5.02	1.817	4.79
Rural	-1.063	-5.61	-1.031	-5.48	-0.194	-5.49	-0.185	-5.25
Sparsity	-0.012	-1.38	-0.012	-1.39	-0.002	-1.19	-0.002	-1.25
Pop1000	-0.026	-1.26	-0.028	-1.34	0.000	0.00	0.000	0.02
Gp1000	0.298	1.80	0.328	1.98	0.048	2.08	0.052	2.24
Pfemale	-0.175	-0.61	-0.070	-0.25	0.026	0.62	0.045	1.10
Avgpage	0.015	1.27	0.012	1.05	0.002	1.12	0.002	0.87
Pcqnonuk	0.146	0.41	0.116	0.32	0.024	0.46	0.027	0.52
Pmscont	0.585	3.02	0.569	2.94	0.045	1.56	0.037	1.26
Dispensing	-0.440	-1.72	-0.500	-1.95	-0.045	-0.94	-0.060	-1.27
Exfh	0.413	3.49	0.473	4.00	0.065	3.42	0.074	3.91
Herf	0.747	2.35	0.640	2.03	0.057	1.07	0.043	0.80
Constant	3.892	3.10	5.793	4.65	-3.072	-16.52	-2.765	-14.99
$\bar{R}^2$	0.5907		0.5919					
Initial Log L					-117311.5		-117311.5	
Model Log L					-102843.9		-102773.0	
N	34705		34705		34705		34705	

[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 40 indicator dummies. Robust t statistics allow for clustering of indicators in practices.
ind upper = 1 if practice above upper threshold in 2005/6 for indicator, 0 otherwise.
lag_ind upper: = 1 if practice above upper threshold in 2004/5 for indicator, 0 otherwise.

In model 1 the coefficient on the dummy *ind upper* (which equals 1 if the practice is above the upper threshold for the indicator in 2005/6) is positive. *ind upper* is positively correlated with unobserved factors which increase exception reporting and thus increase achievement and so may fail to reflect the incentives for practices expecting to be under the upper threshold to increase exceptions.

Model 2 uses *ind upper lag*, a dummy equalling 1 if the practice was above the upper threshold in the previous year. The coefficient on *ind upper lag* is negative and highly significant. It suggests that practices that were above the upper threshold in 2004/5 had exception reporting rates that were about 0.87% lower in 2005/6 than practices which were below the lower threshold in 2004/5. Given that the average exception rate in 2005/6 was 7.92%, the proportionate effect on exception reporting is large.

5. Conclusions

The fact that practices could have treated substantially fewer patients (12.5%) without falling below the upper thresholds for indicators and thereby reducing practice revenue is compatible with altruistic motivation.

Although the median practice achieved 99.2% of the available points, the percentage of patients with a particular condition who received the appropriate care was rather lower at 80.2%.[12]

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). Practices with more female GPs and with younger GPs deliver higher quality.

Reported prevalence was strongly positively correlated with other measures of morbidity and with educational deprivation. Reported prevalence thus appears, unsurprisingly to be strongly affected by genuine prevalence.

Practices with a higher proportion of female GPs or with younger GPs had higher reported prevalence. Since such practices also had higher delivered quality this suggests that female and younger GPs are perhaps more conscientious in delivering care to their patients. The fact that reported prevalence was associated with practice characteristics implies that extreme caution should be exercised in using QOF reported prevalence rates to measure population prevalence of disease.

Our tests show evidence of gaming of prevalence reporting. The lagged measure of the extent to which the practice was above the upper threshold, and hence had an incentive to increase reported prevalence, was positive and statistically significant in models where we pooled observations across disease domains. In the models estimated separately for the eight disease domains, it is positive in all cases and significant in four cases.

Exception reporting is higher in practices in areas with higher rates of population morbidity and with a higher proportion from an ethnic minority. Exception reporting is lower by practices in areas with higher education deprivation and in rural areas. These are factors that plausibly ought to affect exception reporting in the absence of gaming. However, we found that exception reporting was also correlated with practice characteristics, such as whether the practice was an ex-fundholder. Since the permitted grounds for exception reporting relate to patient rather than practice characteristics, this suggests that practices were over reporting exceptions.

Our more direct tests for gaming of exceptions also suggested that there was gaming of exceptions upward for 2005/6 by practices that were below the upper threshold for indicators in 2004/5 and so expected to be otherwise below the upper threshold in 2005/6.

If there are persistent unobserved factors which affect exception reporting then our test for gaming will still be subject to endogeneity bias. However the bias is to increase the coefficient on the lagged dummy variable recording whether the practice was above the upper threshold in 2004/5. The fact that we find negative coefficients that are statistically and quantitatively significant on the lagged upper threshold dummies is strong evidence for gaming of exception reporting.

The introduction of the QOF may have had its intended consequence: the first year of the QOF seems to have show above trend performance against some clinical indicators. However, it also appears to have had unintended consequences in GPs' gaming of reported prevalence and exception rates.

[12] Average over the 41 indicators in Table 5 weighted by the maximum available points.

References

Campbell S, Reeves D, Kontopantelis E, Sibbald B, Roland M. Improvements in clinical quality in English primary care before and after the introduction of a pay for performance scheme: longitudinal cohort study. *New England Journal of Medicine* 2007, in press.

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* 2006; 25: 449-478.

Gulliford M, Ashworth M, Robotham D, Mohiddin A. Achievement of metabolic targets for diabetes in English general practice, 2000 to 2005. *Diabetic Medicine* 2006; in press.

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; in press.

McLean G, Sutton M, Guthrie B. 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* 2006; 60: 917-922.

Morris S, Sutton M, Gravelle H. Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science and Medicine* 2005; 60: 1251-1266.

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

Wagstaff A, van Doorslaer E. Equity in health care finance and delivery. In Culyer, A. and Newhouse, J. (eds.), Handbook of Health Economics 2000, Elsevier, Amsterdam.

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

Practice g revenue from indicator i in disease domain k is

$$\begin{aligned} R_{gki} &= v\left[M_g / \bar{M}\right]\pi_{ki}(\alpha_{gki})\left[P_{gk} / M_g\right]^{\frac{1}{2}}\left[G^{-1}\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-1} \\ &= \bar{v}_g \pi_{ki}(\alpha_{gki}) F_{gk}(P_{gk}) \qquad \text{(A1)} \end{aligned}$$

where $\pi_{gki} = \pi_{ki}(\alpha_{gki}) = \pi_{ki}^o \min\{1, \max\{(\alpha_{gki} - \alpha_{kiL})/(\alpha_{kiU} - \alpha_{kiL}), 0\}\}$ and $\alpha_{gki} = N_{gki}/(P_{gk} - E_{gki})$. For $\alpha_{gki} \in (\alpha_{kiL}, \alpha_{kiU})$, $\partial R_{gki} / \partial P_{gk}$ has the same sign as

$$\begin{aligned} &-\frac{\alpha_{gki}}{(P_{gk} - E_{gki})} P_{gk}^{\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-1} \\ &\quad + (\alpha_{gki} - \alpha_{kiL})\left\{\frac{1}{2} P_{gk}^{-\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-1} - P_{gk}^{\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-2} \frac{1}{2} P_{gk}^{-\frac{1}{2}} M_g^{-\frac{1}{2}}\right\} \\ &< -\frac{\alpha_{gki}}{(P_{gk} - E_{gki})} P_{gk}^{\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-1} + (\alpha_{gki} - \alpha_{kiL}) \frac{1}{2} P_{gk}^{-\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)^{\frac{1}{2}}\right]^{-1} \\ &= \left[-\frac{\alpha_{gki}}{(P_{gk} - E_{gki})} + (\alpha_{gki} - \alpha_{kiL}) \frac{1}{2P_{gk}}\right] P_{gk}^{\frac{1}{2}} \left[\textstyle\sum_h (P_{hk} / M_h)\right]^{-1} < 0 \qquad \text{(A13)} \end{aligned}$$

C. Comparative statics

The practice objective function is

$$s(N,E,P) = bN + R(N,E,P) - C(N,E,P)$$
$$= bN + \bar{v}\pi(\alpha)F(P) \;\; -c_{N1}N - c_{E1}\left(\left|E - E^o\right|/P^o\right) - c_{P1}\left(\left|P - P^o\right|/M\right)$$
$$-\tfrac{1}{2}\left[c_{N2}N^2 + c_{E2}\left((E - E^o)/P^o\right)^2 + c_{P2}\left((P - P^o)/M\right)^2\right] \qquad \text{(A14)}$$

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

$$s_N = b - C_N \le 0, \quad N \ge 0, \quad s_N N = 0 \qquad \text{(A15)}$$
$$s_E^+ = -C_E^+ > 0 > s_E^- = -C_E^- \qquad \text{(A16)}$$
$$s_P^+ = -C_P^+ > 0 > -C_P^- \qquad \text{(A17)}$$

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

$$s_N = b + R_N - C_N = b + \bar{v}\pi'\alpha_N F - C_N = 0 \qquad \text{(A18)}$$
$$s_E = R_E - C_E = \bar{v}\pi'\alpha_E F - C_E = 0 \qquad \text{(A19)}$$
$$s_P = R_P - C_P = \bar{v}\pi'\alpha_P F + \bar{v}\pi(\alpha)F' - C_P = 0 \qquad \text{(A20)}$$

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

$$s_{NN} = -C_{NN} < 0 \qquad \text{(A21)}$$
$$s_{EE} = \bar{v}\pi'\alpha_{EE}F - C_{EE} < 0 \qquad \text{(A22)}$$
$$s_{PP} = \bar{v}\pi'\alpha_{PP}F + 2\bar{v}\pi'\alpha_P F' + \bar{v}\pi F'' - C_{PP} < 0 \qquad \text{(A23)}$$
$$s_{NE} = \bar{v}\pi'\alpha_{NE}F > 0 \qquad \text{(A24)}$$
$$s_{NP} = \bar{v}\pi'\alpha_{NP}F + \bar{v}\pi'\alpha_N F'$$
$$= N^{-1}\left[\bar{v}\frac{\pi^o}{\alpha_U - \alpha_L}\alpha_P F + \bar{v}\frac{\pi^o(\alpha - \alpha_L)}{\alpha_U - \alpha_L}F' + \bar{v}\frac{\pi^o\alpha_L}{\alpha_U - \alpha_L}F'\right]$$
$$= N^{-1}\left[R_P + \bar{v}\frac{\pi^o\alpha_L}{\alpha_U - \alpha_L}F'\right] < 0 \qquad \text{(A25)}$$

$$s_{EP} = \bar{v}\pi'\alpha_{EP}F + \bar{v}\pi'\alpha_E F' = \bar{v}\pi' 2\alpha_{NP}\alpha F + \bar{v}\pi'\alpha_N \alpha F'$$
$$= \alpha\left[s_{NP} + \bar{v}\pi' 2\alpha_{NP}\alpha F\right] < 0 \qquad \text{(A26)}$$

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 (A13) 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

$$\frac{\partial N^{*UL}}{\partial b} = -\left[s_{EE}s_{PP} - (s_{PE})^2\right]\Delta^{-1} > 0 \qquad \text{(A27)}$$

$$\frac{\partial E^{*UL}}{\partial b} = \left[s_{EN}s_{PP} - s_{PN}s_{EP}\right]\Delta^{-1} > 0 \qquad \text{(A28)}$$

$$\frac{\partial P^{*UL}}{\partial b} = -\left[s_{EN}s_{PE} - s_{PN}s_{EE}\right]\Delta^{-1} < 0 \qquad \text{(A29)}$$

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 (A27). The signs (A28) and (A29) 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

$$s_N = b + R_N - C_N = b + \bar{v}\pi'\alpha_N F - C_N = 0 \qquad \text{(A30)}$$

$$s_E^+ = R_E - C_E^+ > 0 > s_E^- = R_E - C_E^- \qquad \text{(A31)}$$

$$s_P = R_P - C_P = \bar{v}\pi'\alpha_P F + \bar{v}\pi(\alpha)F' - C_P = 0 \qquad \text{(A32)}$$

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 (A31) on E is replaced by

$$s_E^+ = R_E - C_E^+ > 0 = s_E^- = R_E - C_E^- \qquad \text{(A33)}$$

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 (A15) 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

$$s_N^+ = b' - C_N = 0 < s_N^- = \bar{v}\pi'(P^o - E^o)^{-1} + b' - C_N \quad \text{(A34)}$$

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

$$b_2 - C_N(N^{*L}(b_2)) = 0 = b_2 + \bar{v}\pi'\left(P^*(b_2) - E^*(b_2)\right)^{-1} - C_N(N^{*LU}(b_2)) \quad \text{(A35)}$$

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

$$s_N = b - C_N = 0 \quad \text{(A36)}$$

$$s_E^+ = -C_E^+ > 0 > -C_E^- \quad \text{(A37)}$$

$$s_P = R_P - C_P = \bar{v}\pi^o F' - C_P = 0 \quad \text{(A38)}$$

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 (A18) on N for solutions with $\alpha \in (\alpha_L, \alpha_U)$ solves for

$$N^{*UL}(b) = \frac{R_N + b - c_{P1}}{c_{P2}} = \frac{\pi^o\left[(\alpha_U - \alpha_L)\left(P^{*UL}(b) - E^{*UL}(b)\right)\right]^{-1} + b - c_{P1}}{c_{P2}} \quad \text{(A39)}$$

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

$$N^{*L}(b) = \frac{b - c_{P1}}{c_{P2}} \quad \text{(A40)}$$

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 (A39) and (A40) 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	+	–