PAPER PRESENTED AT THE HEALTH ECONOMISTS' STUDY GROUP IN YORK, 5-7 JANUARY 2011

Revenue or Reputation – A Quantitative Analysis of General Practitioner Motivation and Peer Effects using Longitudinal Data

Thomas Allen, Matt Sutton
Health Economics, Health Sciences Research Group
University of Manchester
thomas.allen@manchester.ac.uk

Abstract

The Quality and Outcomes Framework (QOF) is a performance rewarding compensation scheme introduced in April 2004 for general practices across the UK. It provides practices with an incentive to improve the level of care they provide by rewarding them with points depending on their achievements on a wide range of indicators. Practices receive income for each point that they achieve, with the revenue per point varying over diseases, practices and time. Practices with higher disease prevalence and more registered patients earn more revenue per point. As they are used for practice comparisons, the points themselves can be seen as a reputational incentive in addition to the financial rewards they invoke. We therefore seek to discover which incentive is strongest; financial or reputational. We find the number of points offered for each patient treated on an indicator to be a significant incentive to increase performance and hence interpret this as a response to reputation. Although revenue incentivises GPs, this is done implicitly, as we find that performance does not change when the revenue offered for each patient treated on an indicator changes. We also investigate how peer effects influence performance on the QOF. We find that an increase in the number of GPs at the practice has a negative effect on aggregate performance and that practices are not influenced by new peers when joining a larger Primary Care Trust. The former is attributed to free-riding and the latter to a decrease in peer effects when the peer group is increased. Our analysis also provides evidence that GPs respond to targets or benchmarks, since their performance is affected by the upper thresholds set on the percentage of patients that must be treated to achieve maximum points.

Introduction

A major component of the new General Medical Services contract introduced in April 2004 was the Quality & Outcomes Framework (QOF). Under the QOF, practices are rewarded for providing a good quality of care to patients and these rewards form one of the main income streams for General Practitioners (GPs). Average net GP incomes rose from £85,000 to £114,000 between 2003/4 and 2005/6, with QOF income representing a third of income in 2004/5 (National Audit Office, 2008).

The QOF has generally been seen as a financial incentive scheme but its origin was in clinical expert opinion of what constituted evidence-based practice (Roland, 2004). The quality scores for individual practices are published and compared between local practices on the NHS Choices website[1]. These scores therefore represent an externally-observed marker of reputation with peers in other practices, with the commissioning organisation (the local Primary Care Trust) and potentially with patients.

We investigate the QOF by focussing on two relatively unexplored areas; peer effects in general practices and how GPs are motivated.

For our motivation analysis we use an unusual feature of the QOF, which has now been abolished, whereby the revenue from activity is neither proportional to the required activity nor to the marker of reputation, the QOF points. The revenue per point varies between diseases within practices, between practices within diseases, and over time in both dimensions. We are thus able to test whether practice effort responds more to the revenue or the reputation payoff using fixed effects regression to control for unobservable heterogeneity in quality between practices.

For our analysis of peer effects we use the number of GPs at the practice as an indicator of within-practice peer effects and the number of practices within the PCT as an indicator of between-practice peer effects. Wang et al (2006) showed that larger practices attained a significantly higher number of QOF points than smaller practices. Larger practices attained a greater number of organisational points but there were no differences in levels of clinical care.

In our analysis we find evidence that peer effects acting within practices are generally negative. We also find that peer effects between practices are eroded when the peer group (PCT) is enlarged. In terms of incentives we find that reputation, rather than income, has an effect on performance.

In the next section we provide an explanation of how the QOF operates. The following section presents our theoretical framework and hypotheses. Details of the data we have used are covered in the next section. The remaining sections cover econometric methods and results. The final section contains our discussion of the results along with policy recommendations and further avenues for research.

[1] http://www.nhs.uk

The QOF

The introduction of the QOF was a significant component of the major overhaul of the GP contract in 2004. An overview of the new scheme is given by Roland (2004). Initially, up to 1050 points were available for practices to earn subject to their achievement on 146 clinical, organisational, patient experience and additional service indicators. More details on the indicators can be found on the NHS Information Centre website[2].

On the clinical indicators, practices are awarded points based on the ratio of the number treated relative to those declared eligible for treatment. Points are awarded in proportion to this ratio between a lower and an upper threshold. Depending on the indicator the upper threshold lies between 55% and 90%. The lower threshold for all indicators was 25% for the first two years and 40% thereafter. Practices maximise revenue by reaching the upper threshold for as many indicators as possible.

For example, the indicator CHD12 was worth 7 points for all years of the QOF but the upper threshold changed from 85% in the first 2 years to 90% for years 3, 4 and 5. This would mean that in 2006/7 a practice would have to treat a higher proportion of patients than it did in the previous years to obtain the same number of points. For full details of the indicators, their points and thresholds see the NHS websites[3] and Table 1.

Achieving points translates into higher revenue. For an average practice each point was worth approximately £75 in the first year and £125 thereafter. However, the level of payment does not depend only on the number of points achieved but also on the practice's disease prevalence rate compared to the national average and the size of the registered population compared to the national average. An illustration of how payments for achievement on clinical indictors are calculated is given in Guthrie et al (2006). Very low prevalence rates are truncated. This truncated prevalence rate is then square rooted and divided by its mean to create an Adjusted Disease Prevalence Factor (ADPF). Once the ADPF is calculated it is multiplied by the ratio of the list size to the mean list size and the nationally-agreed payment per point[4].

The formula for 05/06 onwards is therefore:

$$Payment\ per\ point = \frac{List\ size}{\overline{List\ size}} \times \frac{\sqrt{T.Prevalence}}{\overline{\sqrt{T.Prevalence}}} \times £124.6$$

Where the bar accent indicates a mean value and the "*T.*" indicates truncation.

Practices receive very different levels of payment per point and per patient treated. Guthrie et al (2006) found that income per patient treated varied by as much as 44 times across all practices and between 1.5 and 2.7 times when outlying (top and bottom 5%) practices were excluded. Practices with higher than average prevalence or higher than average list size

[2] http://www.ic.nhs.uk
[3] http://www.nhsemployers.org and http://www.qof.ic.nhs.uk/
[4] £75 or £124.60 depending on year

receive higher amounts per point. The relationships between payment per point and these variables are non-linear. Payment per point varies between practices, between diseases and over time.

Concepts and hypotheses

Financial Incentives

Central to a vast majority of incentive schemes is the desire to solve the *principal-agent* problem; when an agent, who has their own aims, acts on behalf of the principal, with different aims (Lazear & Gibbs 2009). Incentive schemes exist to align the objectives of the agent with those of the principal and there are many theories that seek to explain how best to achieve this. Issues arise as these theories often contradict each other and are hard to empirically test. Contradictions exist, for example, between the predictions based on intrinsic motivation and gift exchange. As Akerlof (1982) explains, a firm wishing to induce more effort from its workers may pay more than the market clearing rate – this is gift exchange. Conversely, paying more has been seen to erode the intrinsic motivation of a task and result in less effort (Ariely 2005).

General support for financial incentives based on performance has come from Lazear (2000). A firm changing from a wage to a piece rate was analysed and the results showed a large significant increase in productivity as a result. This productivity rise was achievable as the output of the workers could be monitored and measured easily. The OQF had made it possible to measure some aspects of a GPs work but there are many areas that cannot to incentivised by a piece rate – their consultant style for example.

More specific support for a compensation scheme based on paying doctors for performance comes from Gaynor and Gertler (1995). Their results suggest that shifting compensation schemes from one not linked to performance to one solely linked to performance would more than double output.

Given the evidence on financial incentives we would hypothesise that the income earned from the QOF would act as a significant incentive to increase output.

Reputation and Altruism

The behaviour of healthcare professionals differs from the typical *principal-agent* model in human resource economics as the healthcare professional acts as an agent to two principals, the patient and the insurer (Ellis and McGuire 1986). The utility function for doctors is typically assumed to demonstrate semi altruism; doctors are concerned with their own income but also with the quality of care they provide (Iversen & Ma 2009) or patient health (Gravelle et al 2002). Incentive schemes for healthcare professionals must therefore be tailored to address these added dimensions.

Siciliani's (2009) formal model of intrinsic motivation assumes that GPs have different levels of altruism and that they derive utility from how they are perceived by others; either generous or greedy. Doctors are assumed to be motivated by a combination of income, patient well-being and their own reputation. Providing a high level of care will have a positive effect on one's reputation. The income earned for a certain task has a negative effect on reputation because doctors may be perceived as greedy if they target high income tasks. Thus, an increase in income may have no overall effect on output or, if the stigma related to perceived greed is great enough, lead to a reduction in output.

Previous work on the motivation coming from reputation suggests that, in addition to the financial incentive to increase output, being perceived as having a better reputation could lead to significant increases in the incentives to improve achievement.

Team Size and Peers

Other contradictions that exist in Human Resources Economics concern team size and peer effects. As the size of a team increases, the result may be a stronger incentive to free-ride and exert less effort (Kandel & Lazear 1992) or, on the other hand, it may cause stronger peer pressure and result in increased effort (Ratto et al, 2001).

Armin and Ichino conducted a controlled field experiment that aimed to find evidence of peer effects. They set up an experiment to compare a single treatment, where the worker was unobserved during their task, and a paired treatment, where workers were able to observe the other's performance. The experiment yielded results in support of peer effects as those working in the paired treatment had significantly higher productivity.

A more general study on the effects of working with peers on output is provided by Kandel and Lazear (1992). They begin by developing the foremost problem associated with working within a team, the free-rider problem, and show that the effort exerted in a team will be lower than the efficient level. This is due to the property of teams to have shared effort and shared reward. Kandel and Lazear show that once peer pressure is added to the workers utility function the equilibrium level of effort is raised; however this peer pressure may have a negative effect on individual utility. Peer pressure is effective only when workers have an incentive *and* the ability to apply such pressure; the former is achieved through profit sharing and the latter from guilt, norms or mutual monitoring (Kandel & Lazear 1992).

Holmström (1982) considers moral hazard and the free-rider problem and concludes that the key areas which an optimal compensation scheme must address are certainty in production, team size and risk aversion. When output is perfectly measurable one should punish the team when output falls below the Pareto Optimum. When output is imperfectly measurable and teams are small, one should punish when output falls below a critical level. When teams are large and consist of risk-averse agents, a first best solution is unattainable and it becomes necessary to measure individual performance. General Practices can be thought of as being small teams and, despite the QOF, still have imperfectly measureable output. If this is the case then a punishment should be given if output falls below a critical level; this level could be the lower threshold.

Gaynor and Gertler (1995) also conclude that a large team would have a lower output per doctor, measured by office visits per week, when compared to a small or average sized team but that this difference diminished as the compensation scale increases.

An in-depth consideration of team based incentives and how these apply to the NHS was performed by Ratto et al (2001). They define the free-rider effect; in terms of shirking, the private marginal cost is shared between the team, so is less than the social marginal cost (Ratto et al, 2001 p.13). Ratto et al also cite Marion & Zabojnik's (2001) suggestion that this problem can be addressed through the use of sub-groups that compete with each other and this can produce results close to optimality when the highest performing team receives a reward. The NHS system of many small and medium sized general practices can be seen as adopting this strategy, though not adopting the relative performance rewards.

When relating the theories of team incentives to the NHS, Ratto et al (2001) emphasise that team production is particularly difficult to identify and define due to the complexities of patient treatment. They highlight some of the benefits and costs of team production. Benefits include; information transfer, mutual monitoring and specialisation. Costs include; free-riding and team conflicts. Whether additional financial incentives will be an effective means of reducing free-riding depends on measuring team performance; if this is accurate then financial incentives may be more effective. They conclude that the decision to use team incentives depends on the size and type of team.

Our analysis of peer effects involves a breakdown of the influences that come from within a particular practice and those which act between practices. To proxy the within practice peer effect we shall use the number of GPs per practice; a proxy is necessary as performance is measured at the practice level. To measure peer effects acting between practices we will take advantage of the change in PCT size that occurred in 2006 and affected around 50% of practices. By calculating the performance of the pre and post 2006 peers we can analyse if new peers had an effect on performance. The literature suggests that peer effects can have either a negative (free-riding) or a positive (peer pressure) effect on performance and therefore a clear hypothesis will not be made at this point.

Data

We use practice level data for clinical indicators and disease prevalence for five years 2004/5-2008/9. For each indicator we have the points awarded, the number of patients treated (numerator) and the number of eligible patients (denominator). From the prevalence dataset we have the percentage prevalence rate for each disease for each practice and the practice list size.

Not all clinical indicators have been consistently included in the QOF over the five-year period. We analyse 23 indicators. A description of these indicators with their points and upper thresholds are shown in Table 1. Summary statistics are given in Table 2.

We include the following additional variables in the analyses: the number of staff per 100,000 of the population at PCT level; the distance to the nearest practice; -the LISI

Deprivation Index; variables which measure the gender/age composition of practice population and a rural/urban dummy variable. Data on the number and gender composition of GPs at each practice were obtained from the NHS Information Centre.

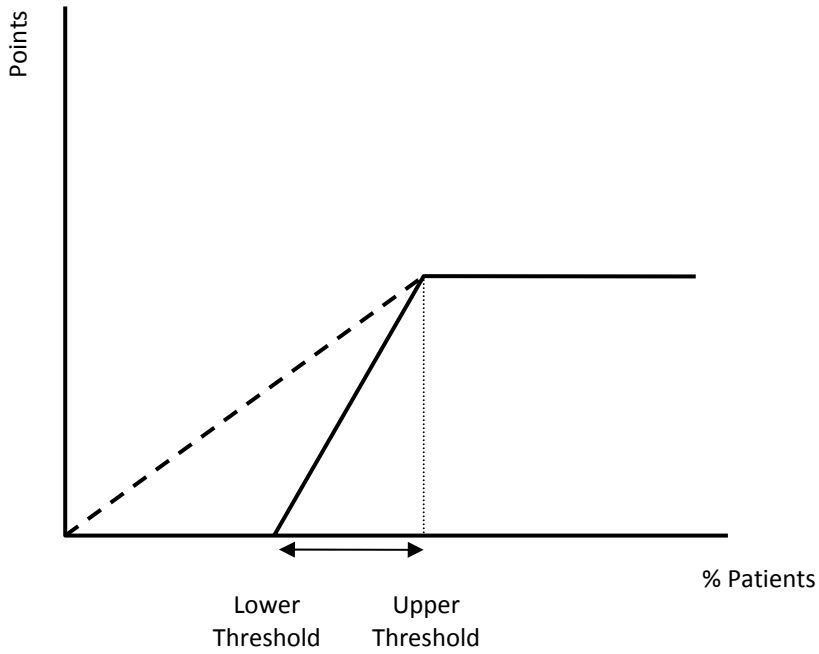
The points per patient for each indicator was defined as the maximum number of points available divided by the number of patients at the practice for which treatment would need to be undertaken to get from the lower threshold to the upper threshold. It is therefore the slope of the points function with respect to the numbers requiring treatment or the marginal points reward per treated patient between the lower and upper threshold.

The second key variable is the revenue earned per indicator per patient. We calculated the total revenue that would be earned by each practice on each indicator if they achieved the upper threshold and divided this by the number of patients requiring treatment to reach the upper threshold from the lower threshold.

The rationale for using the number of patients between the lower and upper thresholds is demonstrated in Figure 1. If the upper threshold was used to calculate the denominator for the two payoff variables (the dotted line in Figure 1) the result would be an inaccurate representation of the marginal payoffs facing most practices. The method used (the solid line) represents the marginal payoff for treating patients between the two thresholds.

This gives two payoff variables for each indicator: additional revenue and additional points earned for each additional patient treated between the lower and upper threshold. Both payoff variables vary across practices, time and indicators because of variations in the number of patients eligible for treatment, the lower and upper thresholds and the points offered for maximum achievement. The revenue payoff additionally varies with prevalence rates and list size.

Figure 1



Method

We estimate models of the form:

$$\frac{\ddot{N}}{D}_{ijt} = \beta_1 \ddot{Rev}_{ijt} + \beta_2 \ddot{Points}_{ijt} + \beta_3 Threshold + \beta_4 \ddot{PCTA}_{it} + \beta_5 \ddot{PCTB}_{it} + \beta_6 d.\ddot{PCTA}_{it} +$$

$$\beta_7 d.\ddot{PCTB}_{it} + \boldsymbol{\beta}\ddot{\boldsymbol{x}}_{\boldsymbol{it}} + \ddot{\varphi}_t + \ddot{\delta}_t + \ddot{\epsilon}_{ijt} \quad \text{t} = 1, \ldots, 5$$

Where N_{ijt} is the number of patients treated (numerator) by practice *i* on indicator *j* at time t, D_{ijt} is the number of eligible patients (denominator), Rev_{ijt} is the associated revenue payoff, $Points_{ijt}$ is the associated points payoff, $\boldsymbol{x}_{it}\boldsymbol{\beta}$ is a vector of covariates, φ_j is an indicator-specific intercept shift, δ_t is a time-specific intercept shift, α_i is a practice-specific time-invariant heterogeneity term and ϵ_{ijt} is an idiosyncratic error term. The elements of $\boldsymbol{x}_{it}$ are variables measuring social deprivation, distance to nearest practice, practice staff per 100,000 of the population, rural/urban dummy, list size, number of GPs per practice, percentage of female GPs, and the age-gender composition of the practice population.

The remaining variables are included to measure between-practice peer effects. For each practice we identify their peers in the pre-reconfiguration set of PCTs (group A) and the new peers they will receive following the reconfiguration of PCTs (group B). In each year we calculate the average points achieved by practices in groups A and B and include these as potential predictors of the practice's performance (variables $PCTA_{it}$and $PCTB_{it}$). The re-configuration occurred in October 2006 so we define the years 2007/8 and 2008/9 as the post-reconfiguration years. We thus interact the $PCTA_{it}$and $PCTB_{it}$ variables with a dummy variable (d) representing the post-reconfiguration period.

Our interpretation of these variables is as follows. The coefficient on the $PCTA_{it}$ variable may reflect peer effects, but it could be endogenous as it reflects the effect of any time-varying unobservable influence on performance that is correlated within PCTs. Our focus is therefore on the interaction term, which shows how the effects of peer performance changes in the post reconfiguration period. The coefficient on the $PCTB_{it}$ variable does not reflect peer effects as practices in group B were not peers to the practice in the pre-reconfiguration period. Its inclusion therefore serves only to reflect spatial correlation in unobservables. Again, the focus is on the interaction term as this reflects the additional effects on performance once the practices in group B become new peers for the practice.

The practice unobservable term allows for any characteristic of the practice that remains constant over time and cannot be included in the model, such as physical or skill features of the practice and the GPs and practice-employed staff.

In addition to models using practices as the grouping variable we will also group observations by practice and indicators as well as practice and diseases. This increases our ability to account for unobserved heterogeneity. Specifically this permits practices to be unobservably better on certain indicators or diseases.

Results

We present our main results in Table 3 which shows our model with the three different groupings of unobservable heterogeneity, as well as a set of results without peer effects or list size. The latter is used to demonstrate that our specification is not affected by list size being used to calculate revenue; to be more precise changes in list size account for some of the variation in our revenue payoff. Therefore, we remove list size, and other variables related to it, and still find an insignificant result meaning the revenue payoff is not unfairly penalised by having some of its variation removed by being picked up in another variable directly.

From the other three specifications presented in Table 3 the point's payoff and the upper threshold variable both have a positive coefficient and considering the scale of the relevant variables these coefficients are strong as well as significant. Our results also reveal that the revenue payoff is insignificant.

A coefficient of -0.44 to -0.47 on Patients per GP suggests a negative relationship between workload and performance. This means that increasing the number of patients per GP will reduce the QOF performance of a practice. The number of GPs measures the intensity of peer effects with our results implying that increasing the number of GPs by 1 will decrease achievement by 0.2%; this result is significant.

Our analysis of between practice peer effects focuses on the two variables that measure the performance of pre and post change PCT and their interactions with the post change period. A positive and significant coefficient on pre and post change PCT performance suggests that practice performance may be affected by the performance of their peers. However, this effect could be due to the endogeneity that exists between practice and PCT performance. This would be the case if there were any time-varying determinants of performance that are spatially correlated. The interaction terms are interpreted by adding their effects to the un-interacted terms. These sums are close to zero for both variables suggesting a removal of the peer effects in the post change period.

Supplementary to the results in Table 3 are those of Tables 4 and 5. Here we show the results of a practice and indicator grouping broken down into the individual disease categories of the data. The purpose of this is to show how the payoff variables fluctuate between different diseases. In Table 5 we provide a complete set of results without omitting any control variables.

Discussion

Since its introduction, the QOF has been the target of a range of studies aimed at uncovering how effective a compensation scheme it represents and if there have been negative or unforeseen consequences. Notwithstanding the amount of analysis conducted with the aid of the QOF, there still exists many questions to which we may find answers within the data. It was the aim of this particular study to approach two relatively unexplored

themes and investigate them further using the QOF. These themes were the within and between practice peer effects and the particular form of motivation which explains the changes in achievement observed since the scheme's introduction.

Our motivation analysis implies that it is both the point's payoff and the value of the upper threshold that are integral in the determination of performance. Both have positive and statistically significant coefficients. On the other hand there is no evidence that practices responded to changes in the revenue payoffs. That is not to say that GPs are not financially motivated, as increased revenue is implied when practices approach the upper threshold and more points are earned. What our result indicates is that the specific way in which the revenue payoff changed during our sampled period, via changes in payments per point, list size or prevalence for example, did not significantly increase the performance of practices. This suggests that practices are motivated to increase performance by points and thresholds and *not* by the explicit revenue that is earned as a result.

From the above we conclude that GPs are motivated by implicit, as opposed to explicit, financial rewards in addition to the reputational rewards of increased points. This conclusion lends support to the Siciliani (2009) model as well as the paying for performance literature such as Lazear (2000). The response GPs are seen to have to upper thresholds is more difficult to place in the literature; this could be another form of reputational or financial motivation.

Focusing now on our peer effects analysis, we conclude from our negative and significant coefficient on the number of GPs per practice that, once we have controlled for GP workload, there are negative peer effects acting within practices. We suggested that negative peer effects can come in the form of free-riding and the evidence now supports this. This is an interesting result as, even though the issue of free-riding is common, one may expect the structure of practices to induce positive peer effects; peer pressure or information transfer for example. However, the QOF rewards are attributed to the practice as a whole and not to individual GPs. This may create the incentive to free-ride as individual performance is often difficult to isolate from team production (Ratto 2001). The strength of this result relies on the strength of the number of GPs as a proxy for peer effects. The ideal measure would be the performance of individual practitioners in each practice but, as QOF performance is at the practice level, this is not possible. It is feasible that there are additional positive effects coming from having high performing peers that are unobservable in our data.

In terms of between practice peer effects we interpret our result as providing mixed evidence for their existence. We saw that pre-2006 peers had a positive, although potentially endogenous, effect on practice performance but this effect is eroded in the post-2006 period. There may be a causal relationship in effect here; positive peer effects are present in smaller PCT but when these practices join larger groupings this peer effect disappears. Post-2006 peers also have a positive effect on performance in the pre-change period, again potentially endogenous, that is unaffected by the organisational change; this is evidenced by the insignificant of the interaction term. In other words joining a larger grouping of practices with a different level of performance has no additional effect on the

practice's performance and practices do not try to raise their performance to catch up with their new peers.

Overall peer effects seem stronger in the smaller PCT. This could be directly due to the PCT size or perhaps the between-practice peer effects operated in the early years of the QOF when the scheme was new and practices were establishing their new reputations. After that point QOF performance increase to a point where there was less pressure coming from within a PCT.

Ratto et al, (2001) suggests that free-riding could be addressed through the use of sub-groups that compete with each other for rewards. However, this is not how the QOF works. The QOF is an absolute performance reward system whereby the performance of one practice does not affect the rewards given to another. An alternative that may address the issue of both between and within peer effects is a relative performance compensation scheme (Ratto et al, 2001). This system may differ from the QOF as practices within a PCT or GPs within a practice could receive additional incentives to be the top performers of their group; this may solve the free-rider problem. It would be difficult to say that a system like this would not be open to increased gaming or cheating due to the presence of even higher powered incentive, it is also likely to reduce cooperation between practices and GPs.

Policy recommendations that arise from this study are four fold. Firstly, the significance of the point's payoff has been interpreted as a result of reputational incentives. We therefore propose an increase in focus on points as a measure of ability/performance to further increase the reputational incentive.

Secondly, given the power of the response to upper thresholds it would seem that adjusting these values would have a substantial effect on motivation and hence performance. If certain indicators are deemed more important than others then raising the upper threshold could have a greater effect than increasing the revenue/points for those indicators.

Thirdly, as a consequence of the insignificance of our revenue payoff the revenue rewards could be standardised payments that did not fluctuated with such things as list size or prevalence. We predict that this would not affect achievement as our evidence suggests that practices are not motivated by the specific revenue per indicator. What this means is if we were to take two indicators identical in every way except that the revenue for one was higher, due to higher prevalence, then the difference in performance should be indiscernible.

Finally, to address the issue of peer effects, a relative performance aspect should be built into the QOF to act across, but not within, practices. This should increase competition while not removing the incentive for GPs within a practice to work together; Iversen & Ma (2009) showed that market forces, such as competition, will influence GP behaviour.

In discussing the results from this analysis we have made both conclusions about the general patterns of motivation and policy recommendations based on these conclusions. When considering our results one should be mindful of their limitations. The analysis suffers from a degree of collinearity between the payoff variables, (0.43), and between the number of GPs and patients per GP, (-0.4), despite our best efforts to remove this. We are also

concerned that our analysis is unable to sufficiently control for regional or practitioner characteristics; for example we would have liked to have known the income per GP before the QOF or their age and experience. There is also an issue involving how the revenue and points payoff variables are generated. The issue is that the number of eligible patients appears in both the denominator of the payoff variables as well as the dependent variable measuring performance.

In future work we hope to isolate which types of GPs respond to the QOF more; perhaps those with lower basic salaries or less experience. In addition it would be interesting to extend our analysis to more clinical as well as organisational indicators. Addressing the problems associated with the generation of the payoff variables would also be the focus of further research.

In this study we had two aims; to determine the role played by peer effects and to determine the motivational force behind GPs performance on the QOF. We have presented evidence to demonstrate that peer effects induce free-riding while reputation seems to be a substantial element of motivation. In addition we have been able to show the motivating effect of increases in the upper thresholds that represent perfect performance.

Table 1 - Indicators

Indicator and description	Points and upper threshold %*†				
Year	1	2	3	4	5
CHD 5. The percentage of patients with coronary heart disease whose notes have a record of blood pressure in the previous 15 months	7-90%	7-90%	7-90%	7-90%	7-90%
CHD 6. The percentage of patients with coronary heart disease, in whom the last blood pressure reading (measured in the last 15 months) is 150/90 or less	19-70%	19-70%	19-70%	19-70%	19-70%
CHD 7. The percentage of patients with coronary heart disease whose notes have a record of total cholesterol in the previous 15 months	7-90%	7-90%	7-90%	7-90%	7-90%
CHD 8.The percentage of patients with coronary heart disease whose last measured total cholesterol (measured in the last 15 months) is 5 mmol/l or less	16-60%	16-60%	**17-70%**	17-70%	17-70%
CHD 9. The percentage of patients with coronary heart disease with a record in the last 15 months that aspirin, an alternative anti-platelet therapy, or an anti-coagulant is being taken	7-90%	7-90%	7-90%	7-90%	7-90%
CHD 10. The percentage of patients with coronary heart disease who are currently treated with a beta blocker (unless a contraindication or side-effects are recorded)	7-50%	7-50%	7-**60%**	7-60%	7-60%
CHD 12. The percentage of patients with coronary heart disease who have a record of influenza vaccination in the preceding 1 September to 31 March	7-85%	7-85%	7-**90%**	7-90%	7-90%
STROKE 5. The percentage of patients with TIA or stroke who have a record of blood pressure in the notes in the preceding 15 months	2-90%	2-90%	2-90%	2-90%	2-90%
STROKE 6. The percentage of patients with a history of TIA or stroke in whom the last blood pressure reading (measured in the last 15 months) is 150/90 or less	5-70%	5-70%	5-70%	5-70%	5-70%
STROKE 7. The percentage of patients with TIA or stroke who have a record of total cholesterol in the last 15 months	2-90%	2-90%	2-90%	2-90%	2-90%
STROKE 8. The percentage of patients with TIA or stroke whose last measured total cholesterol (measured in the last 15 months) is 5mmol/l or less	5-60%	5-60%	5-60%	5-60%	5-60%
STROKE 10. The percentage of patients with TIA or stroke who have had influenza immunisation in the preceding 1 September to 31 March	2-85%	2-85%	2-85%	2-85%	2-85%

Indicator and description	Points and upper threshold				
Year	1	2	3	4	5
Hypertension 4. The percentage of patients with hypertension in which there is a record of the blood pressure in the past 9 months	20-90%	20-90%	20-90%	20-90%	20-90%
Hypertension 5. The percentage of patients with hypertension in whom the last blood pressure (measured in last 9 months) is 150/90 or less	56-70%	56-70%	**57-70%**	57-70%	57-70%
Diabetes 2. The percentage of patients with diabetes whose notes record BMI in the previous 15 months	3-90%	3-90%	3-90%	3-90%	3-90%
Diabetes 5. The percentage of diabetic patients who have a record of HbA1c or equivalent in the previous 15 months	3-90%	3-90%	3-90%	3-90%	3-90%
Diabetes 7. The percentage of patients with diabetes in whom the last HbA1C is 10 or less in last 15 months	11-85%	11-85%	11-**90%**	11-90%	11-90%
Diabetes 11. The percentage of patients with diabetes who have a record of the blood pressure in the past 15 months	3-90%	3-90%	3-90%	3-90%	3-90%
Diabetes 12. The percentage of patients with diabetes in whom the last blood pressure is 145/85 or less	17-55%	17-55%	**18-60%**	18-60%	18-60%
Diabetes 16. The percentage of patients with diabetes who have a record of total cholesterol in the previous 15 months	3-90%	3-90%	3-90%	3-90%	3-90%
Diabetes 17.The percentage of patients with diabetes whose last measured total cholesterol within previous 15 months is 5 or less	6-60%	6-60%	6-**70%**	6-70%	6-70%
Diabetes 18. The percentage of patients with diabetes who have had influenza immunisation in the preceding 1 September to 31 March	3-85%	3-85%	3-85%	3-85%	3-85%
COPD 8. The percentage of patients with COPD who have had influenza immunisation in the preceding 1 September to 31 March	6-85%	6-85%	6-85%	6-85%	6-85%

†All lower thresholds are at 25% for 04/05 and 40% for all other years, *Bold indicates a change in either points or upper threshold in that year

Table 2 - Summary Statistics

Variable[1]	Obs.	Mean	Std. Dev.	Min	Max	Variable	Obs.	Mean	Std. Dev.	Min	Max
diabetes02n	41615	216.34	141.35	0	1566	**stroke06p**	41880	4.94	0.43	0	5
diabetes02d	41615	230.12	149.28	0	1686	**stroke07n**	41880	86.18	68.56	0	595
diabetes02p	41615	2.94	0.25	0	3	**stroke07d**	41880	95.85	75.47	0	641
diabetes05n	41615	222.72	145.90	0	1655	**stroke07p**	41880	1.87	0.30	0	2
diabetes05d	41615	230.83	149.94	0	1697	**stroke08n**	41880	64.29	52.36	0	451
diabetes05p	41615	2.97	0.17	0	3	**stroke08d**	41880	87.78	69.15	0	591
diabetes07n	41615	205.61	136.23	0	1526	**stroke08p**	41880	4.73	0.86	0	5
diabetes07d	41615	224.29	146.60	0	1663	**stroke10n**	41880	77.14	61.70	0	537
diabetes07p	41615	10.76	0.83	0	11	**stroke10d**	41880	87.55	69.96	0	616
diabetes11n	41615	230.14	149.79	0	1675	**stroke10p**	41880	1.91	0.25	0	2
diabetes11d	41615	234.42	152.15	0	1700	**hypertension04n**	41845	720.95	490.90	0	4303
diabetes11p	41615	2.99	0.10	0	3	**hypertension04d**	41845	785.41	533.92	0	4630
diabetes12n	41615	169.98	113.99	0	1484	**hypertension04p**	41845	19.62	1.18	0	20
diabetes12d	41615	220.92	144.37	0	1650	**hypertension05n**	41845	577.19	397.20	0	3668
diabetes12p	41615	17.45	1.22	0	18	**hypertension05d**	41845	755.21	514.58	0	4437
diabetes16n	41615	220.88	144.58	0	1600	**hypertension05p**	41845	54.93	5.56	0	57
diabetes16d	41615	231.61	150.49	0	1688	**CHD Prevalence**	41733	225.47	162.25	0	1733
diabetes16p	41615	2.96	0.19	0	3	**Hypertension Prevalence**	41787	791.77	538.51	0	4675
diabetes17n	41615	172.17	115.84	0	1268	**Diabetes Prevalence**	41740	236.35	150.53	0	1648
diabetes17d	41615	214.52	139.77	0	1488	**COPD Prevalence**	41772	92.14	74.92	0	917
diabetes18n	41615	182.16	118.67	0	1279	**Stroke Prevalence**	41366	101.78	80.39	0	697
diabetes18d	41615	204.01	132.70	0	1464	**CHD Prevalence%**	41787	3.52	1.36	0	63.2
diabetes18p	41615	2.91	0.28	0	3	**Hypertension Prevalence%**	41787	12.32	3.65	0	100
copd08n	41845	73.57	58.90	0	748	**Diabetes Prevalence%**	33558	3.69	1.10	0	16.4
copd08d	41845	80.92	65.64	0	799	**COPD Prevalence%**	41787	1.45	0.84	0	26.3
chd05n	41841	218.25	159.90	0	1688	**Stroke Prevalence%**	41384	1.53	0.86	0	63
chd05d	41841	223.68	163.73	0	1772	**Distance to nearest practice**	41050	1.10	1.98	0	50.88
chd05p	41841	6.97	0.31	0	7	**Practice staff[3]**	40921	64.77	31.52	18.912	196.73
chd06n	41841	191.65	140.73	0	1518	**Rural dummy variable**	40249	0.15	0.36	0	1
chd06d	41841	218.31	159.96	0	1732	**Social deprivation**	939549	12.55	8.95	0.04	90
chd06p	41841	18.91	0.97	0	19	**List size**	41505	6424.37	3972.87	13	39136
chd07n	41841	204.14	150.15	0	1555	**Total Number of GPS**	41495	4.02	2.71	1	29
chd07d	41841	219.23	160.37	0	1737	**Female GPs%**	41495	36.63	27.82	0	100
chd07p	41841	6.83	0.62	0	7	**Numerator (N)[2]**	960742	203.37	229.64	0	4303
chd08n	41841	162.61	121.38	0	1212	**Denominator (D)[2]**	960742	231.84	265.77	0	4630
chd08d	41841	204.83	149.64	0	1597	**Points per patient [2]**	959319	0.27	1.23	0.003	190
chd08p	41841	16.27	1.76	0	17	**Revenue per patient[2]**	883671	17.67	86.99	0.503	21636
chd09n	41841	204.44	150.07	0	1582	**Prop. Male 0-4**	921993	0.03	0.01	0	0.17
chd09d	41841	218.84	160.42	0	1735	**Prop. Female 0-4**	921993	0.03	0.01	0	0.13
chd09p	41841	6.88	0.47	0	7	**Prop. Male 5-14**	921993	0.06	0.01	0	0.50
chd10n	41841	117.82	88.88	0	1004	**Prop. Female 5-14**	921993	0.06	0.01	0	0.15
chd10d	41841	168.98	128.65	0	1424	**Prop. Male 15-44**	921993	0.22	0.05	0	1.00
chd10p	41841	6.85	0.74	0	7	**Prop. Female 15-44**	921993	0.21	0.04	0	0.71
chd12n	41841	181.17	133.06	0	1431	**Prop. Male 45-64**	921993	0.13	0.02	0	1.00
chd12d	41841	199.62	146.79	0	1631	**Prop. Female 45-64**	921993	0.12	0.03	0	0.52
chd12p	41841	6.80	0.63	0	7	**Prop. Male 65-74**	921993	0.04	0.01	0	0.33
stroke05n	41880	96.22	76.05	0	651	**Prop. Female 65-74**	921993	0.04	0.01	0	0.11
stroke05d	41880	99.69	78.73	0	679	**Prop. Male 75-84**	921993	0.02	0.01	0	0.50
stroke05p	41880	1.98	0.14	0	2	**Prop. Female 75-84**	921993	0.03	0.01	0	0.25
stroke06n	41880	82.36	65.53	0	573	**Prop. Male 85+**	921993	0.01	0.01	0	0.39
stroke06d	41880	95.98	75.91	0	655	**Prop. Female 85+**	921993	0.01	0.01	0	0.58

[1] Variables ending in n/d/p refer to Numerator/Denominator/Points achieved respectively, [2] Measured across all indicators, [3] Practice staff per 100,000 of PCT population

Table 3 – Results with different groupings of the fixed effects

	i(practices)	i(practices & diseases)	i(practices & indicators)	i(practices & indicators) No peer effects
Variable	**Coefficient (t-stat)**	**Coefficient (t-stat)**	**Coefficient (t-stat)**	**Coefficient (t-stat)**
Points Payoff	0.57 (3.46) *	0.45 (3.99) *	0.55 (2.61) *	0.85 (5.57) *
Revenue Payoff	-0.00 (-0.56)	0.00 (0.03)	0.00 (0.30)	0.00 (0.51)
Maximum Threshold	0.36 (58.78) *	0.36 (49.77) *	0.35 (52.26) *	0.38 (70.28) *
Number of GPs	-0.13 (-2.74) *	-0.13 (-4.14) *	-0.13 (-6.55) *	-
Patients per GP	-0.47 (3.75) *	-0.45 (-5.68) *	-0.44 (-9.58) *	-
Performance in pre reorganisation PCT	0.02 (5.29) *	0.02 (9.05) *	0.02 (15.08) *	-
Performance in practices that will join post reorganisation	0.01 (2.51) *	0.01 (4.36) *	0.01 (7.18) *	-
Performance in pre reorganisation PCT interaction◆	-0.02 (-4.23) *	-0.02 (-6.61) *	-0.02 (-10.74) *	-
Performance in practices that will join post reorganisation interaction◆	-0.00 (-0.18)	-0.00 (-0.52)	-0.00 (-0.87)	-
Constant	41.27 (5.36) *	32.56 (9.83) *	36.72 (18.09) *	53.11 (45.01) *
Number(observations)	438251	438251	438251	784879
Number(groups)	4309	21534	99055	181222

* Significant at the 95% level

† Included rural dummy, measure of social deprivation, distance to nearest practice, practice staff per 100,000 population, Year dummies, population age-sex composition, % female GPs and indicator dummies.

◆Pre and post change PCT interacted with the post change period (2007/8 and 2008/9)

Table 4 – Individual Disease Results

	CHD	COPD	Stroke	Diabetes	Hypertension
Variable	**Coefficient (t-stat)**	**Coefficient (t-stat)**	**Coefficient (t-stat)**	**Coefficient (t-stat)**	**Coefficient (t-stat)**
Points Payoff	0.88 (3.23) *	3.77 (3.96) *	-0.5 (-0.75)	-1.42 (-4.55) *	0.09 (0.23)
Revenue Payoff	-0.00 (-2.74) *	0.00 (0.10)	0.02 (4.32) *	0.06 (9.53) *	0.01 (1.66)
Maximum Threshold	0.42 (44.23) *	(Dropped)	(Dropped)	0.52 (56.50) *	(Dropped)
Number of GPs	-0.15 (-4.57) *	-0.26 (-2.96) *	-0.11 (-2.23) *	-0.13 (-4.09) *	-0.11 (-2.30) *
Patients per GP	-0.42 (-5.37) *	-0.56 (-2.58) *	-0.48 (-4.16) *	-0.47 (-6.40) *	-0.42 (-3.69) *
Performance in pre reorganisation PCT	0.02 (10.72) *	0.02 (4.64) *	0.01 (4.75) *	0.02 (10.17) *	0.01 (3.47) *
Performance in practices that will join post reorganisation	0.01 (4.48) *	0.01 (1.41)	0.01 (3.71) *	0.01 (3.77) *	0.01 (2.19) *
Performance in pre reorganisation PCT interaction◆	-0.02 (-6.32) *	-0.03 (-4.36) *	-0.02 (-4.62) *	-0.01 (-6.02) *	-0.01 (-3.14) *
Performance in practices that will join post reorganisation interaction◆	-0.00 (-1.86)	-0.01 (-1.67)	0.00 (0.11)	0.003 (1.45)	0.00 (0.11)
Constant	27.95 (8.75) *	72.73 (8.19) *	61.99 (12.37) *	30.46 (9.84) *	58.17 (11.11) *
Number(observations)	143750	20520	101211	131664	41106
Number(groups)	30154	4305	21530	34448	8618

* Significant at the 95% level

† Included rural dummy, measure of social deprivation, distance to nearest practice, practice staff per 100,000 population, Year dummies, population age-sex composition, % female GPs and indicator dummies.

◆ Pre and post change PCT interacted with the post change period (2007/8 and 2008/9)

Table 5 – Results showing all variables, group=practices

Variable	Coefficient (t-stat)	Variable	Coefficient (t-stat)
Points Payoff	0.57 (3.46) *	CHD5 dummy	1.09 (39.06) *
Revenue Payoff	-0.00 (-0.56)	CHD6 dummy	-1.45 (-10.7) *
Maximum Threshold	0.36 (58.78) *	CHD7 dummy	-3.15 (-63.44) *
Number of GPs	-0.13 (-2.74) *	CHD8 dummy	-8.53 (-54.45) *
Patients per GP	-0.47 (-3.75) *	CHD9 dummy	-3.01 (-64.08) *
Performance in pre reorganisation PCT	0.02 (5.29) *	CHD10 dummy	-12.88 (-54.55) *
Performance in practices that will join post	0.01 (2.51) *	CHD12 dummy	-4.42 (-73.72) *
Performance in pre reorganisation PCT interaction◆	-0.01 (-4.23) *	COPD8 dummy	-2.97 (-40.87) *
Performance in practices that will join post reorganisation interaction◆	-0.00 (-0.18)	04/05 dummy	(Dropped)
Distance to nearest practice	(omitted)	05/06 dummy	2.58 (18.48) *
practice staff per 100k	(omitted)	06/07 dummy	4.75 (42.69) *
Rural practice dummy	(omitted)	07/08 dummy	19.63 (5.54) *
Social deprivation	-0.02 (-1.39) *	08/09 dummy	19.99 (5.73) *
Stroke5 dummy	(Dropped)	Prop. Male 0-4	-23.09 (-1.39)
Stroke6 dummy	-3.42 (-24.75) *	Prop. Female 0-4	-20.69 (-1.38)
Stroke7 dummy	0.73 (5.26) *	Prop. Male 5-14	10.91 (0.78)
Stroke8 dummy	-13.00 (64.45) *	Prop. Female 5-14	5.09 (0.37)
Stroke10 dummy	-5.93 (-74.76) *	Prop. Male 15-44	(Dropped)
Hypertension4 dummy	-4.45 (-117.74) *	Prop. Female 15-44	4.43 (0.56)
Hypertension5 dummy	-12.71 (-87.32) *	Prop. Male 45-64	3.27 (0.28)
Diabete2 dummy	-2.22 (41.08) *	Prop. Female 45-64	-7.45 (-0.63)
Diabete5 dummy	0.54 (12.86) *	Prop. Male 65-74	-30.89 (-1.89)
Diabete7 dummy	-3.23 (-55.16) *	Prop. Female 65-74	-8.02 (-0.44)
Diabete11 dummy	1.98 (60.36) *	Prop. Male 75-84	-21.60 (-0.99)
Diabete12 dummy	-8.60 (-37.79) *	Prop. Female 75-84	17.71 (1.06)
Diabete16 dummy	-0.67 (-15.33) *	Prop. Male 85+	-39.57 (-1.12)
Diabete17 dummy	-7.43 (-46.06) *	Prop. Female 85+	10.87 (0.64)
Diabete18 dummy	-4.36 (-56.86) *	Constant	1.71 (0.05)
Number(observations)	438251		
Number(groups)	4309		

* Significant at the 95% level

† Included rural dummy, measure of social deprivation, distance to nearest practice, practice staff per 100,000 population, Year dummies, population age-sex composition, % female GPs and indicator dummies.

◆ Pre and post change PCT interacted with the post change period (2007/8 and 2008/9)

References

Ariely, D. A. N., U. R. I. Gneezy, et al. (2009). "Large Stakes and Big Mistakes." Review of Economic Studies 76(2): 451-469.

Ellis, R. P. and T. G. McGuire (1986). "Provider behavior under prospective reimbursement : Cost sharing and supply." Journal of Health Economics 5(2): 129-151

Falk and Ichino, 2003. Falk, A., Ichino, A., 2003. Clean Evidence on Peer Pressure. CEPR Working Paper DP3834

Gaynor, M. and P. Gertler (1995). "Moral Hazard and Risk Spreading in Partnerships." The RAND Journal of Economics 26(4): 591-613.

Gravelle, H., M. Dusheiko, et al. (2002). "The demand for elective surgery in a public system: time and money prices in the UK National Health Service." Journal of Health Economics 21(3): 423-449.

Guthrie (2006). "Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract." British Journal of General Practice 56: 836-841.

Holmstrom, B. (1982). "Moral Hazard in Teams." The Bell Journal of Economics 13(2): 324-340.

Iversen, Tor and Ma, Ching-to Albert, (2009), Market Conditions and General Practitioners' Referrals, No 2009:8, HERO On line Working Paper Series, Oslo University, Health Economics Research Programme

Kandel, E. and E. P. Lazear (1992). "Peer Pressure and Partnerships." The Journal of Political Economy 100(4): 801-817.

Lazear, E. P. (2000). "Performance Pay and Productivity." The American Economic Review 90(5): 1346-1361.

Lazear, E,. & Gibbs, M,. (2009). "Personnel Economics in Practice". John Wiley & Sons, Inc.

NHS Pay Modernisation: New Contracts for General Practice Services in England, National Audit Office 2008

Ratto, M,. Burgess, S,. Croxson, B,. Jewitt, I,. Propper, C,. Team-Based Incentives in the NHS: an economic analysis. CMPO Working Paper 2001; Series No: 01/37.

Roland, M. (2004). "Linking Physicians' Pay to the Quality of Care — A Major Experiment in the United Kingdom." New England Journal of Medicine 351(14): 1448-1454

Siciliani, L. (2009). "Paying for performance and motivation crowding out." Economics Letters 103(2): 68-71.

Wang YY, O'Donnell CA, Mackay DF, Watt GCM. Practice size and quality attainment under the new GMS contract: a cross-sectional analysis. British Journal of General Practice. 2006;56:830–835